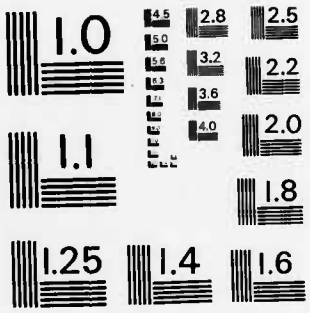
AD-A132 303 ENRICHED MULTINORMAL PRIORS REVISITED(U) CALIFORNIA UNIV BERKELEY OPERATIONS RESEARCH CENTER W S JEWELL NOV 82 ORC-82-14 AFOSR-81-0122 1/1

UNCLASSIFIED F/G 12/1 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

ADA132303

(10)

ENRICHED MULTINORMAL PRIORS REVISITED

by

William S. Jewell
Department of Industrial Engineering
and Operations Research
University of California, Berkeley

DTIC
ELECTE
SEP 12 1983
S D
B

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

DTIC FILE COPY

NOVEMBER 1982 ORC 82-14

This research was supported by the Air Force Office of Scientific Research (AFSC), USAF, under Grant AFOSR-81-0122 with the University of California. Reproduction in whole or in part is permitted for any purpose of the United States Government.

83 09 07 009

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ORC 82-14	2. GOVT ACCESSION NO. AD A132 303	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) ENRICHED MULTINORMAL PRIORS REVISITED		5. TYPE OF REPORT & PERIOD COVERED Research Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) William S. Jewell		8. CONTRACT OR GRANT NUMBER(s) AFOSR-81-0122
9. PERFORMING ORGANIZATION NAME AND ADDRESS Operations Research Center University of California Berkeley, California 94720		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 2304/A5 61102F
11. CONTROLLING OFFICE NAME AND ADDRESS United States Air Force Air Force Office of Scientific Research Bolling Air Force Base, D.C. 20332		12. REPORT DATE November 1982
		13. NUMBER OF PAGES 50
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Multivariate Exponential Family Distributions
Natural Conjugate Priors
Normal, Wishart, Student-t Distributions
Credibility Theory
Bayesian Prediction

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

(SEE ABSTRACT)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ABSTRACT

In a 1974 paper, the author indicated how natural conjugate priors for multidimensional exponential family likelihoods could be enriched in certain cases through linear transformations of independent marginal priors. In particular, it was shown how the usual Normal-Wishart prior for the multinormal distribution with unknown mean vector and precision matrix could have the number of hyperparameters increased; the "thinness" of the traditional prior is well-known. The new, linearly-dependent prior leads to full-dimensional credibility prediction formulae for the observational mean vector and covariance matrix, as contrasted with the simpler, self-dimensional forecasts obtained in prior literature. However, there was an error made in the sufficient-statistics term of the covariance predictor which is corrected in this work. In addition, this paper explains in detail the properties of the enriched multinormal prior and why revised statistics are needed, and interprets the important relationship between the linear transformation matrix and the matrix of credibility time constants. An enumeration of the additional number of hyperparameters needed for the enriched prior shows its value in modelling multinormal problems; it is shown that the estimation of these hyperparameters can be carried out in a natural way, in the space of the observable variables.

Accession For
NTIS GRA&I ✓
DTIC TAB ☐
Unannounced ☐
Justification
By
Distribution/
Availability Codes
Dist | Avail and/or Special
A

DTIC COPY INSPECTED 2

ENRICHED MULTINORMAL PRIORS REVISITED

by

William S. Jewell

1. INTRODUCTION

The multivariate Normal distribution continues to occupy a central role in Bayesian analysis, not only because it is the assumed or the limiting distribution in many practical models, but also because it provides direct access to the study of covariance between random variables. If both the mean and the covariance are random parameters, then the usual informative prior that is assumed is the Normal-Wishart, due to Ando and Kaufman (1965). However, it is well-known that this prior is too "thin", that is, only has a small number of hyperparameters (see, e.g., Press (1981)); this in turn limits the modelling of prior experience. As we shall see, a similar problem occurs in finding a natural conjugate prior for any distribution from the multivariate exponential family.

In (1974b), the author suggested a method for "enriching" the multivariate prior through linear transformations on independent marginals, thus introducing more hyperparameters. This approach successfully enriched the "credibility formula" for the vector mean of the Bayesian predictive distribution, that was also too thin with the Normal-Wishart prior. However, because the original article was in a European actuarial journal, it received little attention, and it is hoped that this paper will encourage further work on this difficult problem. We also take this opportunity to correct several typographic errors and two erroneous formulae in the original work.

We begin with a brief review of the credibility theory that motivated this research and the problem of finding natural conjugate priors for multivariate distributions.

2. CREDIBILITY THEORY

Credibility theory is the name given by American actuaries to an approximate formula developed in the 1920's to forecast the mean of future observations (of insurance claims). In modern terminology, we would say that the problem is one of finding the best linear-least-squares approximation to the mean of the Bayesian predictive density. By now, the literature of credibility theory has grown rapidly; convenient articles showing the variety of models are: Norberg (1979), Kahn (ed.) (1975), and De Vylder (1982). Of course, this theory has many results in common with traditional statistics and with linear filter theory, see, e.g., Diaconis and Ylvisaker (1979) and Feinberg (1980).

To illustrate the basic model development that motivated the work on enriched priors, consider the usual Bayesian set-up in which a p-dimensional vector of random variables*, $\tilde{\underset{\sim}{x}}$, defined over some *fixed* space X in R^p depends upon an abstract (vector or scalar) parameter $\tilde{\theta}$ in a space Θ through a *likelihood density* $p(\underset{\sim}{x} \mid \theta)$; $\tilde{\theta}$ is considered also to be random, with a known *prior-parameter density* $p(\theta)$, developed through previous experience or personal belief. In actuarial science, as in economics, there are few philosophical barriers to such assumptions. If n independent samples of $\tilde{\underset{\sim}{x}}$ are observed with fixed θ , then the posterior-to-data density of $\tilde{\theta}$ is given by Bayes' law:

$$p(\theta \mid \mathcal{D}_x) \propto \prod_{t=1}^{n} p(\underset{\sim}{x}_t \mid \theta)p(\theta) \tag{1}$$

*Notational remarks and distributional results are given in the Appendix.

(we omit the normalization), where $\mathcal{D}_x$ is the data set $\{\underset{\sim}{x}_t \mid t = 1, \ldots, n\}$. In insurance application, the focus of interest is not on θ (which represents some abstract property of the insured risk), but rather on the forecast of some *future outcome* of $\tilde{\underset{\sim}{x}}$, given that θ remains constant (i.e., on the distribution of future claims from the same insured). Thus, attention shifts from (1) to the *predictive density*:

$$p(\underset{\sim}{x} \mid \mathcal{D}_x) = \int p(\underset{\sim}{x} \mid \theta) p(\theta \mid \mathcal{D}_x) d\theta \ . \tag{2}$$

Ideally, the actuary would prefer to express his prior experience not through $p(\theta)$, but through the *prior-outcome (marginal) density:*

$$p(\underset{\sim}{x}) = \int p(\underset{\sim}{x} \mid \theta) p(\theta) d\theta \ . \tag{3}$$

However, it will be seen that is not possible to completely avoid consideration of the structure of $p(\theta)$.

The mean of (2) is the "fair future premium" that is of central interest in insurance, and in the 1920's, American actuaries developed a one-dimensional approximation formula through practical arguments, which we may write:

$$\begin{gathered} E(\tilde{x} \mid \mathcal{D}_x) \approx (1 - z_{00})m + z_{00}\bar{x} \\ z_{00} = \frac{n}{n_{00} + n} \ . \end{gathered} \tag{4}$$

Here, $m = E(\tilde{x})$ is the *mean prior outcome* ("the manual fair premium"), and $\bar{x} = \frac{1}{n} \sum x_t$ is the usual *sample mean* ("the experience fair premium"). z_{00} is the "credibility" factor which mixes the predictor from prior experience, m, with the predictor from experimental evidence, $\bar{x}$; $z_{00} \to 1$ as $n \to \infty$.

n_{00} , the time constant of learning from experience, was set empirically. This credibility method of "experience rating" has worked well in the insurance industry since that time.

With the renewal of interest in Bayesian formulae in the '50s, Bailey and Mayerson showed that (4) was, in fact, *exact* for certain simple conjugate prior $p(x \mid \theta)$ and $p(\theta)$. Bühlmann (1967) then proved the important result that, for *arbitrary* one-dimensional priors and likelihoods, (4) was the best linear least-squares approximation, provided that the time constant n_{00} is chosen to be:

$$n_{00} = EV\{\tilde{x} \mid \tilde{\theta}\}/VE\{\tilde{x} \mid \tilde{\theta}\} . \tag{5}$$

These moments are recognizable as the components of the total variance $V\{\tilde{x}\}$ of (3), and show to what extent the structure of inter-risk and intra-risk variability needs to be specified, a priori.

In (1974a), the author showed that (4) was also exact for the *linear exponential family*, whose likelihood can be written:

$$p(x \mid \theta) = \frac{a(x)e^{-\theta x}}{c(\theta)} , \quad (x \in X) \tag{6}$$

provided that the *natural conjugate prior density*,

$$p(\theta) = \frac{[c(\theta)]^{-n_{00}} e^{-\theta x_0}}{d(\theta)} , \quad (\theta \in \Theta) \tag{7}$$

is used over the complete parameter space Θ for which $d(\theta)$ is finite, and provided that $p(\theta) \to 0$ at both ends of the range (Jewell, 1975a). It turns out that n_{00} is just the time constant n_{00} of Bühlmann.

With these useful results relating the one-dimensional credibility approximation with corresponding Bayesian formulae for which (4) is exact, it was natural to consider extensions to higher dimensions. In an unpublished 1973 report, the author showed that the p-dimensional analogue to the Bühlmann result is:

$$E(\tilde{\underset{\sim}{x}} \mid \mathcal{D}_x) \approx (\underset{\sim}{I} - \underset{\sim}{Z})\underset{\sim}{m} + \underset{\sim}{Z}\tilde{\underset{\sim}{x}} , \tag{8}$$

where now the linear approximation minimizes the *sum* of squared-errors in each dimension. $\bar{\underset{\sim}{x}}$ is the vector sample mean:

$$\bar{\underset{\sim}{x}} = \frac{1}{n} \sum \underset{\sim}{x}_t , \tag{9}$$

and $\underset{\sim}{I}$ is the p-dimensional unit matrix. To develop the remaining first and second moments, we define the conditional vector mean and conditional matrix covariance using $p(\underset{\sim}{x} \mid \theta)$:

$$\underset{\sim}{m}(\theta) = E\{\tilde{\underset{\sim}{x}} \mid \theta\} \quad ; \quad \underset{\sim}{C}(\theta) = \underset{\sim}{V}\{\tilde{\underset{\sim}{x}} \mid \theta\} , \tag{10}$$

and obtain the three unconditional first and second moments:

$$\begin{aligned} \underset{\sim}{m} &= E\underset{\sim}{m}(\tilde{\theta}) \quad ; \quad \underset{\sim}{D} = \underset{\sim}{V}\{\underset{\sim}{m}(\tilde{\theta})\} ; \\ \underset{\sim}{E} &= E\{\underset{\sim}{C}(\tilde{\theta})\} \quad ; \quad \underset{\sim}{C} = \underset{\sim}{D} + \underset{\sim}{E} = \underset{\sim}{V}\{\tilde{\underset{\sim}{x}}\} . \end{aligned} \tag{11}$$

$\underset{\sim}{E}$ and $\underset{\sim}{D}$ are sometimes called the intra-risk and inter-risk components of variance, respectively. As in the one-dimensional case, the analyst must estimate not only $\underset{\sim}{m}$ and $\underset{\sim}{C}$ in X , but must know enough about the parameterization to estimate the two components of covariance.

(8) is seen to again be a mixture of the mean prior outcome, $\underset{\sim}{m}$, and the MLE experience estimator, $\bar{\underset{\sim}{x}}$. However, the weights of the mixture now come from a $p \times p$ *matrix credibility factor*, $\underset{\sim}{Z}$, so that the experience in *all* dimensions is (usually) useful in forecasting any particular $\tilde{x}_i$. In this case, we say we have *full-dimensional credibility*; such forecasts are intuitively better than just forecasting $\tilde{x}_i$ using $(\bar{\underset{\sim}{x}})_i$.

If we define the *matrix of time constants*:

$$\underset{\sim}{N} = \underset{\sim}{E}\underset{\sim}{D}^{-1} , \tag{12}$$

it turns out that the matrix credibility factor is analogous to (4):

$$\underset{\sim}{Z} = n(\underset{\sim}{N} + n\underset{\sim}{I})^{-1} \quad ; \quad (\underset{\sim}{I} - \underset{\sim}{Z}) = \frac{1}{n}(\underset{\sim}{Z}\underset{\sim}{N}) = \frac{1}{n}(\underset{\sim}{N}\underset{\sim}{Z}) .$$

$\underset{\sim}{Z}$, $(\underset{\sim}{I} - \underset{\sim}{Z})$, $\underset{\sim}{N}$, and their powers and inverses all commute, even though $\underset{\sim}{N}$ is not necessarily symmetric. Furthermore, if we find the eigenvalues $\{\nu_i\}$ of $\underset{\sim}{N}$ through $|\underset{\sim}{N} - \nu\underset{\sim}{I}| = 0$, then the eigenvalues of $\underset{\sim}{Z}$ are $\{n/(n + \nu_i)\}$. One can use this to show that, in the non-degenerate case, $\underset{\sim}{Z} \to \underset{\sim}{I}$ as $n \to \infty$, so that the classical estimator $\bar{\underset{\sim}{x}}$ is the ultimate credibility forecast. However, without further restrictions, it is possible for the components of $\underset{\sim}{Z}$, which are rational functions of n , to show non-monotone behavior.

With this "shrinkage" result established, the author next sought to find multi-dimensional priors and likelihoods for which (8) is exact. But here certain difficulties arose that will be explained after we consider multi-dimensional exponential families.

3. NATURAL CONJUGATE PRIORS IN THE MULTIVARIATE EXPONENTIAL FAMILY

The essential reasons why (6) (7) provide the exact Bayesian result (4) is that $\bar{x}$ is the sufficient statistic for θ, because (6) is in the exponential family, and (7) is the natural conjugate prior to (6), which is *closed under sampling*, that is, the *posterior density* $p(\theta \mid \mathcal{D}_x)$ is in the same family as (7), with the updating $n_{00} \leftarrow n_{00} + n$, and $x_0 \leftarrow x_0 + n\bar{x}$. We now attempt to generalize this approach.

If $\tilde{\underset{\sim}{x}}$ is a p-dimensional random variable which depends upon a q-dimensional vector of parameters $\underset{\sim}{\theta}$, the general *multivariate exponential likelihood* can be written:

$$p(\underset{\sim}{x} \mid \underset{\sim}{\theta}) = \frac{a(\underset{\sim}{x})e^{-\underset{\sim}{\theta}'\underset{\sim}{f}(\underset{\sim}{x})}}{c(\underset{\sim}{\theta})} \quad , \quad (\underset{\sim}{x} \in X) \tag{13}$$

where a is the *kernel*, $\underset{\sim}{f}$ is the vector-valued function of *sufficient statistics*, and c is the *normalization factor*, chosen so that $\iiint p(\underset{\sim}{x} \mid \underset{\sim}{\theta})d\underset{\sim}{x} = 1$. The parameter space Θ consists of all points in R^q for which c is finite; Θ is known to be convex but not much else is known in general.

Natural conjugate priors for random $\tilde{\underset{\sim}{\theta}}$ have been constructed for several specific multivariate distributions in the exponential family (see, e.g., Johnson and Kotz (1972)); however, there seems to be little discussion in the literature about how to proceed in general. Based upon one-dimensional procedures, the usual approach is to regard (13) as a function of $\underset{\sim}{\theta}$, and to replace functions of $\underset{\sim}{x}$ and any constants by *hyperparameters*; this gives a prior density:

$$p(\underset{\sim}{\theta}) = p(\underset{\sim}{\theta} \mid n_{00}, \underset{\sim}{f}_{00}) = \frac{[c(\underset{\sim}{\theta})]^{-n_{00}} e^{-\underset{\sim}{\theta}'\underset{\sim}{f}_{00}}}{d(n_{00}, \underset{\sim}{f}_{00})} \quad , \quad (\underset{\sim}{\theta} \in \Theta) \tag{14}$$

with $q + 1$ hyperparameters $(n_{00}, \underset{\sim}{f}_{00})$ that may have to be restricted so that the normalization, d, exists. Usually (14) leads to a prior of recognizable and "interesting" form, even though no advance assurances can be given in the general case. However, if $p(\underset{\sim}{\theta})$ is found to be an "honest" density, then we see immediately that it is closed under sampling; i.e., given n independent samples $(\underset{\sim}{x}_t \; ; \; t = 1, \ldots, n)$ from (13) (with fixed $\underset{\sim}{\theta}$), we find $p(\underset{\sim}{\theta} \mid \mathcal{D}_x)$ to be of form (14), with updating;

$$n_{00} \leftarrow n_{00} + n \quad ; \quad \underset{\sim}{f}_{00} \leftarrow \underset{\sim}{f}_{00} + \sum_{t=1}^{n} \underset{\sim}{f}(\underset{\sim}{x}_t) \; . \tag{15}$$

It is now also clear why $\underset{\sim}{f}$ is called the vector of sufficient statistics.

Most applications of interest concern predictive distributions (Aitchinson and Dunsmore (1975)). A priori, the marginal distribution of $\tilde{\underset{\sim}{x}}$ is:

$$p(\underset{\sim}{x}) = \int p(\underset{\sim}{x} \mid \underset{\sim}{\theta}) p(\underset{\sim}{\theta}) d\underset{\sim}{\theta} = a(\underset{\sim}{x}) \frac{d(n_{00} + 1 \, ; \underset{\sim}{f}_{00} + \underset{\sim}{f}(\underset{\sim}{x}))}{d(n_{00}; \underset{\sim}{f}_{00})} \; , \tag{16}$$

which is also usually of "interesting" form, if (13) and (14) are. Furthermore, it follows that the predictive density $p(\underset{\sim}{x} \mid \mathcal{D}_x)$ is also closed under sampling, and uses the updating (15) in both numerator and denominator of (16).

In Jewell (1974b), it was shown that a certain generalized credibility forecast of the mean value of the function $\underset{\sim}{f}$ then follows:

$$E\{\underset{\sim}{f}(\tilde{\underset{\sim}{x}}) \mid \mathcal{D}_x\} = (1 - z_{00}) E\{\underset{\sim}{f}(\tilde{\underset{\sim}{x}})\} + z_{00} \frac{1}{n} \sum_{t=1}^{n} \underset{\sim}{f}(\underset{\sim}{x}_t) \; , \tag{17}$$

with a *scalar* credibility factor equal to (4). This result requires rather

strong assumptions about the regularity of $p(\underset{\sim}{\theta})$ on the boundary of Θ, which, however, usually seem to be satisfied for practical distributions.

If we specialize to the *linear exponential family* with $\underset{\sim}{f}(\underset{\sim}{x}) = \underset{\sim}{x}$ (and $q = p$), we find the exact multi-dimensional mean forecast:

$$E(\tilde{\underset{\sim}{x}} \mid \mathcal{D}_x) = (1 - z_{00})\underset{\sim}{m} + z_{00}\bar{\underset{\sim}{x}} , \tag{18}$$

which should be compared with the credibility approximation (8). This result is obtained, for example, if $\tilde{\underset{\sim}{x}}$ is multinormal, with a random mean vector $\tilde{\underset{\sim}{\mu}}$ and fixed precision matrix $\underset{\sim}{W}$, with a multinormal prior on $\tilde{\underset{\sim}{\mu}}$:

$$\begin{aligned} (\tilde{\underset{\sim}{x}} \mid \tilde{\underset{\sim}{\mu}}) &\sim N_p(\tilde{\underset{\sim}{\mu}};\underset{\sim}{W}) \\ \tilde{\underset{\sim}{\mu}} &\sim N_p(\underset{\sim}{m};n_{00}W) \end{aligned} \tag{19}$$

(see Appendix for distributions). However, because z_{00} is a scalar credibility factor, we see that (18) is rather uninteresting compared to (8), since the forecast of each $\tilde{x}_i$ is given by a credibility forecast using only $(\bar{\underset{\sim}{x}})_i$. Furthermore, since $\underset{\sim}{N} = n_{00}\underset{\sim}{I}$, each forecast has the *same* time constant n_{00}! We shall call such forecasts *self-dimensional*, in contrast to the *full-dimensional* form (8). To rectify this unsatisfactory state of affairs, we shall have to consider ways to "enrich" (18), by adding more hyperparameters to (14) or (19).

4. ENRICHED PRIORS FROM LINEARLY DEPENDENT MULTIVARIATE EXPONENTIAL FAMILIES

In Jewell (1974b), the observation was made that the thinness of (14) was due to the scalar nature of the factor $[c(\underset{\sim}{\theta})]^{-n_{00}}$.

There have been various proposals for enriching multivariate priors, particularly the multinormal prior. Those known to the author are:

(1) Press (1981) refers first to enrichment as the process of examining the behavior of $p(\underset{\sim}{x} \mid \underset{\sim}{\theta})$ as a function of $\underset{\sim}{\theta}$, which we used to find (14), above. But he also suggests that various lower-dimensional marginals, considered as functions of $\underset{\sim}{\theta}$, could be multiplied together to get $p(\underset{\sim}{\theta})$. Such a procedure was followed by Kaufman for the multinormal priors in unpublished papers in 1965 and 1967; however, it seems to lead to rather complex results with little motivation, and does not solve the full-dimensional credibility problem;

(2) One can, of course, also multiply (14) by an arbitrary function, say $[g(\underset{\sim}{\theta})]^{-n_1}$. But, since n_1 will not participate in the updating (15) because $g(\underset{\sim}{\theta})$ has no relation to $c(\underset{\sim}{\theta})$, it is essentially a nuisance hyperparameter, and no real enrichment has taken place, except perhaps to put the prior in some standard form. See also the scale changes in Section 10;

(3) Another procedure, often recommended when one-dimensional priors do not match empirical priors, is to use a *model-mixture prior*, by combining a (small) number of natural conjugate priors (14) with *different coefficients*:

$$p(\underset{\sim}{\theta}) = \sum_k \alpha_k p(\underset{\sim}{\theta} \mid n_{k0}, \underset{\sim}{f}_{k0}) \ . \tag{20}$$

Here the $\left\{\alpha_k \geq 0 \ ; \ \sum \alpha_k = 1\right\}$ and the various hyperparameters are determined by matching the empirical prior. The difficulty with this approach is not with the updating, which follows (15) for every "model" in (20), but is that the weighting coefficients, $\alpha_k \leftarrow \alpha_k(\mathcal{D}_x)$, *must also be updated*, usually

in a messy algebraic manner. And, of course, when applied to the credibility problem, (20) will merely stabilize, for large n, on the "best" n_{i0} for self-dimensional forecasts;

(4) Another suggestion for enrichment is to regard (14) as part of an *hierarchical model* (Jewell (1975b)), with the hyperparameters $(n_{00}, \underset{\sim}{f}_{00})$ also considered as random variables, with their own hyperprior density $p(n_{00}, \underset{\sim}{f}_0)$. This approach requires an extremely simple normalization d, and probably extremely few random hyperparameters in order to carry out the necessary marginalization. Dickey, Lindley, Press and James, (1981) do this with a model in which there are only 2 random and 1 fixed hyperparameters.

Since the thinness of (14) is due to the scalar nature of the factor $[c(\underset{\sim}{\theta})]^{-n_{00}}$, it follows that the number of time-constant hyperparameters can be increased if $c(\underset{\sim}{\theta})$ can be decomposed into several factors that depend upon various subsets of $\underset{\sim}{\theta}$. One way the decomposition might occur was if both the sufficient-statistics vector, $\underset{\sim}{f}(\underset{\sim}{x})$, and the kernel, $a(\underset{\sim}{x})$ could be decomposed into related components, through a *linear transformation*; it follows then that $c(\underset{\sim}{\theta})$ would undergo a similar decomposition, and each factor could have its own hyperparameter, n_{i0}. We call such likelihoods where this decomposition works *linearly dependent multivariate exponential families*. (LDMEFs).

To fix ideas, consider first the case where the sample mean is the only sufficient statistic, i.e., $\underset{\sim}{f}(\underset{\sim}{x}) = \underset{\sim}{x}$ and $q = p$. Let $\underset{\sim}{A}$ be an invertible $p \times p$ matrix, and consider first a vector $\underset{\sim}{y}$ of *independent* random risks, each component having a linear exponential likelihood with parameters $(\phi_1, \phi_2, \ldots, \phi_p)$, so that the joint likelihood density is:

$$p(\underset{\sim}{y} \mid \underset{\sim}{\phi}) = \prod_{i=1}^{p} \frac{b_i(y_i)}{c_i(\phi_i)} e^{-\underset{\sim}{\phi}'\underset{\sim}{y}} \quad , \quad (\underset{\sim}{y} \in \mathcal{Y}) \tag{21}$$

for some appropriate vector space, $\mathcal{Y}$, kernels $\{b_i\}$, and normalizations $\{d_i\}$. For each component i , an appropriate natural conjugate prior similar to (7) is constructed, so that the (independently distributed) $\{\tilde{\phi}_i\}$ have a joint prior:

$$p(\underset{\sim}{\phi}) = \prod_{i=1}^{p} \frac{[c_i(\phi_i)]^{-n_{i0}}}{d_i(n_{i0}, y_{i0})} e^{-\underset{\sim}{\phi}'\underset{\sim}{y}_0} \quad , \quad (\underset{\sim}{\phi} \in \Phi) \tag{22}$$

for some appropriate vector space, Φ , and the $2p$ hyperparameters, $\underset{\sim}{y}_0' = [y_{10}, y_{20}, \ldots, y_{p0}]$ and $\underset{\sim}{n}_0' = [n_{10}, n_{20}, \ldots, n_{p0}]$. Already the situation is somewhat improved, for it can be seen that the exact form for $E\{\tilde{\underset{\sim}{y}} \mid \mathcal{D}_y\}$, although "self-dimensional" as in (18), now has a *different time constant*, n_{i0} , for each dimension. In other words, the matrix credibility forecast (8) is exact, but with a diagonal matrix of time constants:

$$\underset{\sim}{N}_0 = \text{diag}\ (\underset{\sim}{n}_0) = \begin{pmatrix} n_{10} & & & 0 \\ & n_{20} & & \\ & & \ddots & \\ 0 & & & n_{p0} \end{pmatrix} , \tag{23}$$

and corresponding diagonal credibility matrices $\underset{\sim}{Z}_0 = n(\underset{\sim}{N}_0 + n\underset{\sim}{I})^{-1}$.

Now, make the (full rank) transformations:

$$\underset{\sim}{y} = \underset{\sim}{A}^{-1}\underset{\sim}{x} \quad ; \quad \underset{\sim}{\phi} = \underset{\sim}{A}'\underset{\sim}{\theta} \; ; \tag{24}$$

this changes (21) into a likelihood proportional to:

$$p(\underset{\sim}{x} \mid \underset{\sim}{\theta}) \propto \prod_{i=1}^{p} b_i\left((\underset{\sim}{A}^{-1}\underset{\sim}{x})_i\right) e^{-\underset{\sim}{\theta}'\underset{\sim}{x}} \; , \tag{25}$$

and by comparison with (13), with $\underset{\sim}{f}(\underset{\sim}{x}) = \underset{\sim}{x}$, we see that this is precisely the result we will have *if* the linear transformation $\underset{\sim}{x} = \underset{\sim}{A}\underset{\sim}{y}$ factors the kernel $b(\underset{\sim}{x})$ into p components, each depending upon a *single* y_i $(i = 1,2, \ldots, p)$.

If the likelihood permits this factorization, it then follows that the *enriched LDMEF prior* on $\tilde{\underset{\sim}{\theta}}$ is gotten from (22) as:

$$p(\underset{\sim}{\theta}) \propto \prod_{i=1}^{p} [c_i((\underset{\sim}{A}'\underset{\sim}{\theta})_i)]^{-n_{i0}} e^{-\underset{\sim}{\theta}'\underset{\sim}{x}_0} \; , \tag{26}$$

with the hyperparameter transformation

$$\underset{\sim}{x}_0 = \underset{\sim}{A}\underset{\sim}{y}_0 \; . \tag{27}$$

The following result is then obtained:

<u>Theorem</u>:

If $p(\underset{\sim}{x} \mid \underset{\sim}{\theta})$ is an LDMEF likelihood, as defined in (21) (25), and the enriched prior (26) is used with hyperparameters $\{n_{i0} > 0\}$, the full-dimensional forecast (8) for $E\{\tilde{\underset{\sim}{x}} \mid \mathcal{D}_x\}$ applies, with time constant matrix:

$$\underset{\sim}{N} = \underset{\sim}{A}\underset{\sim}{N}_0\underset{\sim}{A}^{-1} = \underset{\sim}{E}\underset{\sim}{D}^{-1} \; . \tag{28}$$

Proof:

As mentioned above, the proof requires the assumption that $p(\underset{\sim}{\theta})$ is well-behaved on the boundary of Θ. After showing that $\underset{\sim}{x}_0 = \underset{\sim}{N}\underset{\sim}{m}$, the result follows after showing that the correct updating with data $\mathcal{D}_x$ is given by:

$$\underset{\sim}{N} \leftarrow \underset{\sim}{N} + n\underset{\sim}{I} \quad ; \quad \underset{\sim}{x}_0 \leftarrow \underset{\sim}{x}_0 + \sum_{t=1}^{n} \underset{\sim}{x}_t \ . \tag{29}$$

Further details may be found in Jewell (1974b). ■

To show that this result is not vacuous, we reconsider the problem of the multinormal prior with random mean $\tilde{\underset{\sim}{\mu}}$ and fixed precision matrix $\underset{\sim}{W}$, whose thin prior was given in (19). It follows from the above that the "thick" prior that gives a full-dimensional credibility forecast is simply:

$$\tilde{\underset{\sim}{\mu}} \sim N_p(\underset{\sim}{m} \ ; \ \underset{\sim}{W}\underset{\sim}{N} = \underset{\sim}{N}'\underset{\sim}{W}) \ . \tag{30}$$

Here $\underset{\sim}{E} = \underset{\sim}{W}^{-1}$ is fixed, but $\underset{\sim}{D} = (\underset{\sim}{W}\underset{\sim}{N})^{-1} = \underset{\sim}{N}^{-1}\underset{\sim}{E}$. The initial precision matrix associated with the forecast $E\{\tilde{\underset{\sim}{x}} \mid \mathcal{D}_x\} = E\{\underset{\sim}{\mu} \mid \mathcal{D}_x\}$ is $\underset{\sim}{N}\underset{\sim}{W} = \underset{\sim}{D}^{-1}$, so that, after n observations, the precision improves (homoscedastically) to $(\underset{\sim}{N} + n\underset{\sim}{I})\underset{\sim}{W} = \underset{\sim}{D}^{-1} + n\underset{\sim}{E}^{-1}$. These results are well-known (see, inter alia, De Groot (1970), page 175).

We turn now to the fundamentally more difficult problem of the multinormal with both random mean and random precision.

5. THE ANDO-KAUFMAN MULTINORMAL PRIOR

Consider the p-dimensional multinormal with random mean vector $\underset{\sim}{\mu}$, and random precision matrix $\underset{\sim}{\Omega}$, for which the likelihood density is:

$$p(\underset{\sim}{x} \mid \underset{\sim}{\mu},\underset{\sim}{\Omega}) = (2\pi)^{-p/2}|\underset{\sim}{\Omega}|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}(\underset{\sim}{x}-\underset{\sim}{\mu})'\underset{\sim}{\Omega}(\underset{\sim}{x}-\underset{\sim}{\mu})\right\} . \tag{31}$$

This is in the family (13), with $f_i(\underset{\sim}{x}) = \underset{\sim}{x}_i$ and $\underset{\sim}{\theta}_i = -(\underset{\sim}{\Omega}\underset{\sim}{\mu})_i$ for $i = 1,2, \ldots, p$, and $f_i(\underset{\sim}{x}) = \frac{1}{2} x_j x_k$ and $\theta_i = \omega_{jk}$ for $(j,k) = 1,2, \ldots, p$, and $i = k + pj = p + 1, p + 2, \ldots, (p + p^2) = q$. Thus, we have many more random parameters than observables. The kernel $a(\underset{\sim}{x}) \equiv 1$, and the normalizing factor (in traditional notation) is:

$$c(\underset{\sim}{\mu},\underset{\sim}{\Omega}) = (2\pi)^{p/2}|\underset{\sim}{\Omega}|^{-\frac{1}{2}} \exp\left\{\frac{1}{2}\underset{\sim}{\mu}'\underset{\sim}{\Omega}\underset{\sim}{\mu}\right\} . \tag{32}$$

To form the natural conjugate prior, we first assume $q + 1 = p^2 + p + 1$ hyperparameters $\{n_{00};\underset{\sim}{x}_0;\underset{\sim}{Q}_0\}$, where $\underset{\sim}{Q}_0$ is a $p \times p$ matrix, and then follow (13), with (32) expressed in terms of $\underset{\sim}{\theta}$. Upon transforming back to traditional notation (the Jacobian of the transformation is $|\underset{\sim}{\Omega}|$), we obtain the thin prior:

$$p(\underset{\sim}{\mu};\underset{\sim}{\Omega}) \propto |\underset{\sim}{\Omega}|^{\frac{1}{2}(n_{00}+2)} \exp\left\{-\frac{1}{2}\underset{\sim}{\mu}'(n_{00}\underset{\sim}{\Omega})\underset{\sim}{\mu} + \underset{\sim}{\mu}'\underset{\sim}{\Omega}\underset{\sim}{x}_0 - \frac{1}{2}\operatorname{tr}(\underset{\sim}{\Omega}\underset{\sim}{Q}_0)\right\} . \tag{33}$$

By factorization into $p(\underset{\sim}{\mu} \mid \underset{\sim}{\Omega}) \cdot p(\underset{\sim}{\Omega})$, this can be seen to be a *Normal-Wishart density*. For the conditional mean:

$$p(\underset{\sim}{\mu} \mid \underset{\sim}{\Omega}) \propto |\underset{\sim}{\Omega}|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\underset{\sim}{\mu} - n_{00}^{-1}\underset{\sim}{x}_0\right)'(n_{00}\underset{\sim}{\Omega})\left(\underset{\sim}{\mu} - n_{00}^{-1}\underset{\sim}{x}_0\right)\right\} , \tag{34}$$

that is, a multinormal, with $(\tilde{\underset{\sim}{\mu}} \mid \underset{\sim}{\Omega}) \sim N_p\left(n_{00}^{-1}\underset{\sim}{x}_0; n_{00}\underset{\sim}{\Omega}\right)$.

This leaves, after some algebra, the precision matrix density,

$$p(\underset{\sim}{\Omega}) \propto |\underset{\sim}{\Omega}|^{\frac{1}{2}(n_{00}+1)} \exp\left\{-\frac{1}{2}\,\mathrm{tr}\,(\underset{\sim}{\Omega}\underset{\sim}{U}_0)\right\} , \tag{35}$$

with

$$\underset{\sim}{U}_0 = \underset{\sim}{Q}_0 - n_{00}^{-1}\underset{\sim}{x}_0\underset{\sim}{x}_0' , \tag{36}$$

defined only in those points of $R^{\frac{1}{2}p(p+1)}$ for which $\tilde{\underset{\sim}{\Omega}}$ is positive definite. (35) is the *Wishart density*, $\tilde{\underset{\sim}{\Omega}} \sim W(n_{00} + p + 2 \; ; \; \underset{\sim}{U}_0)$, for which moments are given in the Appendix. We now note that $\underset{\sim}{U}_0$ (and hence $\underset{\sim}{Q}_0$) *must be symmetric*, not for the trace, but for the moment formulae (39) below. Thus, there are really only $\frac{1}{2}(p^2 + 3p + 2)$ free hyperparameters.

The marginal distribution of $\underset{\sim}{\mu}$ requires some additional algebra (see De Groot (1970) or Press (1981)), which finally gives:

$$p(\underset{\sim}{\mu}) \propto \left[1 + n_{00}\left(\underset{\sim}{\mu} - n_{00}^{-1}\underset{\sim}{x}_0\right)'\underset{\sim}{U}_0^{-1}\left(\underset{\sim}{\mu} - n_{00}^{-1}\underset{\sim}{x}_0\right)\right]^{-\frac{1}{2}(n_{00}+p+3)} , \tag{37}$$

which is seen to be a multivariate Student-t density, $\tilde{\underset{\sim}{\mu}} \sim S_p\left(n_{00} + 3 \; ; \; n_{00}^{-1}\underset{\sim}{x}_0 \; ; \; n_{00}(n_{00} + 3)\underset{\sim}{U}_0^{-1}\right)$. Again, moments are found in the Appendix.

Similarly, the marginal outcome density is also found to be Student-t, with $\tilde{\underset{\sim}{x}} \sim S_p\left(n_{00} + 3 \; ; \; n_{00}^{-1}\underset{\sim}{x}_0 \; ; \; n_{00}(n_{00} + 3)(n_{00} + 1)^{-1}\underset{\sim}{U}_0^{-1}\right)$. The updating is, from (31):

$$n_{00} \leftarrow n_{00} + n \quad ; \quad \underset{\sim}{x}_0 \leftarrow \underset{\sim}{x}_0 + \sum_{t=1}^{n} \underset{\sim}{x}_t \quad ; \quad \underset{\sim}{Q}_0 \leftarrow \underset{\sim}{Q}_0 + \sum_{t=1}^{n} \underset{\sim}{x}_t\underset{\sim}{x}_t' . \tag{38}$$

The relationship with the moments defined in (11) follows from the moments of (35) (37):

$$\underset{\sim}{m} = n_{00}^{-1}\underset{\sim}{x}_0 \quad ; \quad \underset{\sim}{E} = (n_{00} + 1)^{-1}\underset{\sim}{U}_0 \quad ; \quad \underset{\sim}{D} = n_{00}^{-1}(n_{00} + 1)^{-1}\underset{\sim}{U}_0 \quad ; \quad \underset{\sim}{C} = n_{00}^{-1}\underset{\sim}{U}_0 \ , \tag{39}$$

so that, as expected, the forecast mean $E(\tilde{\underset{\sim}{x}} \mid \mathcal{D}_x)$ is of the self-dimensional form (18). A new credibility formula of interest is the forecast of the covariance matrix of (37), as updated:

$$\begin{aligned} \underset{\sim}{V}\{\tilde{\underset{\sim}{x}} \mid \mathcal{D}_x\} = (1 - z_{00})\underset{\sim}{C} + z_{00}\left[\frac{1}{n}\sum_{t=1}^{n}(\underset{\sim}{x}_t - \bar{\underset{\sim}{x}})(\underset{\sim}{x}_t - \bar{\underset{\sim}{x}})'\right] \\ + z_{00}(1 - z_{00})(\bar{\underset{\sim}{x}} - \underset{\sim}{m})(\bar{\underset{\sim}{x}} - \underset{\sim}{m})' \ . \end{aligned} \tag{40}$$

Again we see the familiar convex combination of the prior outcome covariance and the classical MLE covariance estimator, supplemented here by an intermediate term which uses the variation of the sample means about their true values $(z_{00}(1 - z_{00})$ attains its maximum value at $n = n_{00})$. Results analogous to (40) are given by both De Groot (1970) and Press (1981), although not in as appealing a form.

The prior (33) was discovered by Ando and Kaufman (1965), and its "thinness" is well-known. The usual criticism is that one cannot set both the means and covariances of $\underset{\sim}{\mu}$ and $\underset{\sim}{\Omega}$ independently, or to put it differently, once $E\{\underset{\sim}{\mu}\}$ and $E\{\tilde{\underset{\sim}{\Omega}}\}$ or $E\{\tilde{\underset{\sim}{\Omega}}^{-1}\}$ are given, there is only one free parameter. From our point of view, the limitation is that the two components of observational covariance cannot be specified independently, since $\underset{\sim}{E} = n_{00}\underset{\sim}{D}$, and this makes the credibility forecasts of mean and covariance (18) (40) both self-dimensional.

Actually, the prior given by Ando and Kaufman is slightly more general than (33), with $|\underset{\sim}{\Omega}|^{\frac{1}{2}(n_{00}+2)}$ replaced by $|\underset{\sim}{\Omega}|^{\frac{1}{2}(\alpha-p)}$, where α is the "degrees of freedom", not necessarily equal to $p + n_{00} + 2$. This leads to invariant nuisance hyperparameter enrichment, of the type already discussed, that merely scales the observational covariances, independent of the means. As mentioned above, in some unpublished work in 1965 and 1967, Kaufman additionally enriched this prior by multiplying the Wishart density by arbitrary powers of the products of determinants of principal minors of $\tilde{\underset{\sim}{\Omega}}$, thus introducing $p - 1$ additional hyperparameters. But, the resulting formulae are quite complicated, and do not appear to give credibility results.

6. LINEARLY-DEPENDENT MULTINORMAL PRIOR

We now use the methods of Section 4 to provide a LDMEF prior to (31), keeping in mind that we want to obtain a full-dimensional forecast (8) and, hopefully, to generalize (40). If we apply the transformations:

$$\tilde{\underset{\sim}{x}} = \underset{\sim}{A}\tilde{\underset{\sim}{y}} \quad ; \quad \tilde{\underset{\sim}{\mu}} = \underset{\sim}{A}\tilde{\underset{\sim}{\lambda}} \; ; \tag{41}$$

we see that we obtain transformed variables $(\tilde{\underset{\sim}{y}};\tilde{\underset{\sim}{\lambda}})$ of full rank, since $\underset{\sim}{A}$ is assumed to be $p \times p$ and invertible. But this then leads in (31) to a transformation of the precision matrix:

$$\tilde{\underset{\sim}{\Pi}} = \underset{\sim}{A}'\tilde{\underset{\sim}{\Omega}}\underset{\sim}{A} \quad ; \quad \tilde{\underset{\sim}{\Omega}} = (\underset{\sim}{A}^{-1})'\tilde{\underset{\sim}{\Pi}}\underset{\sim}{A}^{-1} \; ; \tag{42}$$

and, since we require that the $\{\tilde{y}_i\}$ be statistically independent, the *transformed precision matrix* $\tilde{\underset{\sim}{\Pi}}$ *must be diagonal with probability one*! In other words, to factor the last p^2 components of $\{f_i(\underset{\sim}{x})\}$, we must impose constraints on the associated parameters. This also factors the term $|\underset{\sim}{\Omega}|^{\frac{1}{2}}$.

This then permits one to introduce a random mean, $\tilde{\lambda}_i$, and a random precision, $\tilde{\pi}_i$, for each $\tilde{y}_i$, and to set $(\tilde{y}_i \mid \lambda_i,\pi_i) \sim N_1(\lambda_i,\pi_i)$, so that the equivalent of (21) is:

$$p(\tilde{\underset{\sim}{y}} \mid \underset{\sim}{\lambda},\underset{\sim}{\Pi}) \propto |\underset{\sim}{\Pi}|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}(\underset{\sim}{y} - \underset{\sim}{\lambda})'\underset{\sim}{\Pi}(\underset{\sim}{y} - \underset{\sim}{\lambda})\right\} , \tag{43}$$

with $\underset{\sim}{\lambda}' = [\lambda_1, \ldots, \lambda_p]$; $\underset{\sim}{\pi}' = [\pi_1,\pi_2, \ldots, \pi_p]$; and $\underset{\sim}{\Pi} = \text{diag}\,(\underset{\sim}{\pi})$.

Through inverse transformations, (43) then reverts to (31), although we must be careful in the sequel with terms involving $\tilde{\underset{\sim}{\Omega}}$, since it is no longer of rank $\frac{1}{2}p(p + 1)$. If preferred, one can think of a full-rank $\tilde{\underset{\sim}{\Omega}}$ being constrained by $\frac{1}{2}p(p - 1)$ equations taken from (42):

$$\sum_k \sum_l a_{ki} a_{lj} \tilde{\omega}_{kl} = 0 \qquad (\text{w.p.1}\ ,\ i \neq j)\ . \tag{44}$$

This loss of rank will be somewhat compensated for by the introduction of more hyperparameters.

The independent natural-conjugate prior for the random parameters $(\tilde{\lambda}_i, \tilde{\pi}_i)$ turns out to be a one-dimensional version of (33):

$$\begin{aligned} p(\lambda_i, \pi_i) &\propto \pi_i^{\frac{1}{2}(n_{i0}+2)} \exp\left\{-\frac{1}{2}\lambda_i^2(n_{i0}\pi_i) + \lambda_i \pi_i y_{i0} - \frac{1}{2} r_{i0}\pi_i\right\} \\ &\propto \left[\pi_i^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\lambda_i - n_{i0}^{-1} y_{i0}\right)^2 (n_{i0}\pi_i)\right\}\right] \\ &\quad \left[\pi_i^{\frac{1}{2}(n_{i0}+1)} \exp\left\{-\frac{1}{2} v_{i0}\pi_i\right\}\right], \end{aligned} \tag{45}$$

which can be called a Normal-Gamma, since $(\tilde{\lambda}_i \mid \pi_i) \sim N_1\left(n_{i0}^{-1} y_{i0}; n_{i0}\pi_i\right)$ and $(\tilde{\pi}_i) \sim G\left(\frac{1}{2}(n_{i0}+3)\ ;\ \frac{1}{2} v_{i0}\right)$, where $v_{i0} = r_{i0} - n_{i0}^{-1} y_{i0}^2$. The marginal density of $\tilde{\lambda}_i$ is a one-dimensional Student-t, similar to (37):

$$p(\lambda_i) \propto \left[1 + n_{i0} v_{i0}^{-1}\left(\lambda_i - n_{i0}^{-1} y_{i0}\right)^2\right]^{-\frac{1}{2}[n_{i0}+4]} . \tag{46}$$

For completeness, we record the first and second moments of $(\tilde{\lambda}_i \mid \pi_i)$:

$$E\{\tilde{\lambda}_i \mid \pi_i\} = n_{i0}^{-1} y_{i0} \quad ; \quad V\{\tilde{\lambda}_i \mid \pi_i\} = (n_{i0}\pi_i)^{-1} \tag{47}$$

and of $\tilde{\pi}_i$ and $\tilde{\pi}_i^{-1}$;

$$E\{\tilde{\pi}_i\} = (n_{i0} + 3)v_{i0}^{-1} \quad ; \quad E\{\tilde{\pi}_i^{-1}\} = (n_{i0} + 1)^{-1}v_{i0} \; ;$$
$$V\{\tilde{\pi}_i\} = 2(n_{i0} + 3)v_{i0}^{-2} \quad ; \quad V\{\tilde{\pi}_i^{-1}\} = 2(n_{i0} + 1)^{-2}(n_{i0} - 1)^{-1}v_{i0}^2 \; ; \tag{48}$$

so that, unconditioning:

$$E\{\tilde{\lambda}_i\} = n_{i0}^{-1}y_{i0} \quad ; \quad V\{\tilde{\lambda}_i\} = n_{i0}^{-1}(n_{i0} + 1)^{-1}v_{i0} \; ; \tag{49}$$

and

$$E\{\tilde{y}_i\} = n_{i0}^{-1}y_{i0} \quad ; \quad V\{\tilde{y}_i\} = n_{i0}^{-1}v_{i0} \; . \tag{50}$$

These are analogous to the Ando-Kaufman results (39), and show that, to obtain meaningful densities, we must pick $y_{i0} > 0$, $v_{i0} > 0$, and $n_{i0} > 0$ ($n_{i0} > 1$ if we want $VV\{\tilde{y}_i \mid \tilde{\pi}_i\}$ to exist). Further, since $v_{i0} = r_{i0} - n_{i0}^{-1}y_{i0}^2$, this imposes a restriction of $r_{i0}n_{i0} > y_{i0}^2$ in the joint prior (45). In fact, since there are only three hyperparameters (n_{i0}, y_{i0}, r_{i0}) , we see that there is already some "thinness" in specifying prior moments in one dimension.

Updating with independent data $\mathcal{D}_y = \{\underset{\sim}{y}_t \; ; \; t = 1,2, \ldots, n\}$ is simply:

$$n_{i0} \leftarrow n_{i0} + n \; ;$$
$$y_{i0} \leftarrow y_{i0} + \sum_{t=1}^{n} y_{it} \; ; \tag{51}$$
$$r_{i0} \leftarrow r_{i0} + \sum_{t=1}^{n} (y_{it})^2 \; .$$

Our next step is to express the joint density of all $(\tilde{\lambda}_i, \tilde{\pi}_i)$ in matrix notation by using the previous definitions for $\underset{\sim}{n}_0$, $\underset{\sim}{N}_0$, and $\underset{\sim}{y}_0$ from Section 4, and setting

$$\underset{\sim}{r}_0' = [r_{10}, r_{10}, r_{p0}] \quad ; \quad \underset{\sim}{R}_0 = \text{diag}\,(\underset{\sim}{r}_0) \; . \tag{52}$$

We then obtain:

$$p(\underset{\sim}{\lambda}, \underset{\sim}{\pi}) \propto \prod_{i=1}^{p} (\pi_i)^{\frac{1}{2}(n_{i0}+2)} \exp\left\{-\frac{1}{2}\underset{\sim}{\lambda}'(\underset{\sim}{\Pi}\underset{\sim}{N}_0)\underset{\sim}{\lambda} + \underset{\sim}{\lambda}'\underset{\sim}{\Pi}\underset{\sim}{y}_0 - \frac{1}{2}\underset{\sim}{\pi}'\underset{\sim}{r}_0\right\} . \tag{53}$$

We can write the last term as $\underset{\sim}{\pi}'\underset{\sim}{r}_0 = \text{tr}\,(\underset{\sim}{\Pi}\underset{\sim}{R}_p)$ if we remember that, with $\underset{\sim}{\Pi}$ diagonal, the trace "annihilates" any off-diagonal elements, so that $\text{tr}\,(\underset{\sim}{\Pi}\underset{\sim}{R}_1) = \underset{\sim}{\pi}'\underset{\sim}{r}_0$ for *any* $\underset{\sim}{R}_1$ with diagonal equal to $\underset{\sim}{r}_0$! This point was responsible for several errors in the author's original paper. We now make the transformations $\underset{\sim}{\lambda} = \underset{\sim}{A}^{-1}\underset{\sim}{\mu}$, $\underset{\sim}{\Pi} = \underset{\sim}{A}'\underset{\sim}{\Omega}\underset{\sim}{A}$, and define:

$$\begin{aligned} \underset{\sim}{N}_0 = \underset{\sim}{A}^{-1}\underset{\sim}{N}\underset{\sim}{A} = \underset{\sim}{A}'\underset{\sim}{N}'(\underset{\sim}{A}')^{-1} \quad &; \quad \underset{\sim}{R}_0 = \underset{\sim}{A}^{-1}\underset{\sim}{Q}_0(\underset{\sim}{A}^{-1})' \; ; \\ \underset{\sim}{N} = \underset{\sim}{A}\underset{\sim}{N}_0\underset{\sim}{A}^{-1} \quad &; \quad \underset{\sim}{Q}_0 = \underset{\sim}{A}\underset{\sim}{R}_0\underset{\sim}{A}' \; ; \end{aligned} \tag{54}$$

and thus obtain the new, enriched LDMEF multinormal prior:

$$p(\underset{\sim}{\mu}, \underset{\sim}{\Omega}) \propto \prod_{i=1}^{p} [(\underset{\sim}{A}'\underset{\sim}{\Omega}\underset{\sim}{A})_i]^{\frac{1}{2}(n_{i0}+2)} \exp\left\{-\frac{1}{2}\underset{\sim}{\mu}'(\underset{\sim}{\Omega}\underset{\sim}{N})\underset{\sim}{\mu} + \underset{\sim}{\mu}'\underset{\sim}{\Omega}\underset{\sim}{x}_0 - \frac{1}{2}\underset{\sim}{1}'\underset{\sim}{A}^{-1}\underset{\sim}{\Omega}\underset{\sim}{Q}_0\underset{\sim}{A}\underset{\sim}{1}\right\} \tag{55}$$

where $\underset{\sim}{1}$ is the $p \times 1$ vector with all 1's $(\underset{\sim}{\Pi} \cdot \underset{\sim}{1} = \underset{\sim}{\pi} \; ; \; \underset{\sim}{R}_0 \cdot \underset{\sim}{1} = \underset{\sim}{r}_0)$. $\underset{\sim}{Q}_0$ is symmetric by construction, but $\underset{\sim}{N}$ need not be.

By comparison with the thin prior (33), we see that $(n_{00}\underset{\sim}{\Omega})$ in the exponent is replaced by $(\underset{\sim}{\Omega}\underset{\sim}{N})$, a generalization similar to the constant covariance case (30). On the other hand, it does not seem possible to simplify the leading product terms in (55) in terms of $\underset{\sim}{N}$, unless, of course, all the time constants are the same, and $\underset{\sim}{N}_0 = n_{00}\underset{\sim}{I} = \underset{\sim}{N}$ transforms the leading product into $|\underset{\sim}{A}|^2|\underset{\sim}{\Omega}|^{\frac{1}{2}(n_{00}+2)}$.

The messy last term in the exponent can be rewritten $\mathrm{tr}\,[\underset{\sim}{A}^{-1}\underset{\sim}{\Omega}\underset{\sim}{Q}_0\underset{\sim}{A}\underset{\sim}{1}\underset{\sim}{1}']$, or simply $\mathrm{tr}\,(\underset{\sim}{\Omega}\underset{\sim}{Q}_0)$ *if* we remember the previous caution about the trace, and realize we would get the same answer now for $\mathrm{tr}\,[\underset{\sim}{\Omega}\underset{\sim}{Q}_1]$, where $\underset{\sim}{Q}_1 = \underset{\sim}{A}\underset{\sim}{R}_1\underset{\sim}{A}'$, if $\underset{\sim}{R}_1$ were *any* square matrix with diagonal equal to $\underset{\sim}{r}_0$.

In fact, because the same matrix $\underset{\sim}{A}$ has been used to transform the original 3p hyperparameters in $(\underset{\sim}{y}_0,\underset{\sim}{N}_0,\underset{\sim}{R}_0)$, it follows that there are strong relations among the $p + p^2 + \frac{1}{2}(p^2 + p) = \frac{3}{2}(p^2 + p)$ transformed hyperparameters $(\underset{\sim}{x}_0;\underset{\sim}{N};\underset{\sim}{Q}_0)$, and the linearly-dependent random parameters $(\underset{\sim}{\mu},\underset{\sim}{\Omega})$, such as:

$$\underset{\sim}{N}\underset{\sim}{Q}_0 = \underset{\sim}{Q}_0\underset{\sim}{N}' \; ; \tag{56}$$

(not expected as $\underset{\sim}{N}$ assymmetric), and:

$$\underset{\sim}{\tilde{\Omega}}\underset{\sim}{N} = \underset{\sim}{N}'\underset{\sim}{\tilde{\Omega}} \quad ; \quad \underset{\sim}{\tilde{\Omega}}\underset{\sim}{Q}_0 = (\underset{\sim}{A}\underset{\sim}{A}')^{-1}(\underset{\sim}{Q}_0\underset{\sim}{\tilde{\Omega}})(\underset{\sim}{A}\underset{\sim}{A}') \quad ; \quad (\text{w.p.1}) \tag{57}$$

which are alternate versions of the $\frac{1}{2}(p^2 - p)$ constraints (44). In short, we must *not* assume that the $\underset{\sim}{\tilde{\Omega}}$, $\underset{\sim}{N}$, and $\underset{\sim}{Q}_0$ of (55) have all the same properties as $\underset{\sim}{\tilde{\Omega}}$, $n_{00}\underset{\sim}{I}$, and $\underset{\sim}{Q}_0$ of (33) or $\underset{\sim}{N}$ of (12). We will return to this point later.

7. LDMEF MULTINORMAL PRIOR MARGINAL AND MOMENTS

As tempting as it is to factor (55) directly, as in Section 5, we are on safer ground if we factor (53) directly, and then transform; by checking also the transformation of moments, this will enable us to deduce the correct form for the new degenerate covariance term.

For the conditional mean, it is easily seen that $(\tilde{\lambda} \mid \Pi) \sim N_p\left(N_0^{-1} y_0; \Pi N_0\right)$, which transforms into $(\tilde{\mu} \mid \Omega) \sim N_p\left(N^{-1} x_0; \Omega N\right)$. This is the natural generalization of (34), and is the same as in the fixed covariance case (30). The moments corresponding to (47) are easily found to be:

$$E\{\tilde{\mu} \mid \Omega\} = N^{-1} x_0 \quad ; \quad V\{\tilde{\mu} \mid \Omega\} = (\Omega N)^{-1} . \tag{58}$$

It is the first term that gives a full-dimensional credibility forecast; note also that the constraint (57) on $\tilde{\Omega}$ guarantees also symmetry of the covariance term.

The difficulty comes from the marginal precisions, which have joint density:

$$p(\pi) \propto \prod_{i=1}^{p} (\pi_i)^{\frac{1}{2}(n_{i0}+1)} \exp\left\{-\frac{1}{2} \pi' v_0\right\} , \tag{59}$$

where v_0 was previously defined to have components $r_{i0} - n_{i0}^{-1} y_{i0}^2$. We know that we want to express at least the exponent in terms of the matrix Π before transformation; r_0 has a diagonal version, but it is not clear how to diagonalize y_{i0}^2, as we want the updating of x_0 to remain simple. Two possibilities are to define a matrix version of v_0 as:

$$V_1 = R_0 - N_0^{-1} y_0 y_0' \quad or \quad V_2 = R_0 - y_0 y_0' N_0^{-1} . \tag{60}$$

($\underset{\sim}{V}_1$ was used in the author's 1974 paper). Both of these are easily transformed, and both give the correct result under annihilation by the trace operator, $\mathrm{tr}\ (\underset{\sim}{\Pi}\underset{\sim}{V}_1) = \mathrm{tr}\ (\underset{\sim}{\Pi}\underset{\sim}{V}_2) = \underset{\sim}{\pi}'\underset{\sim}{v}_0$.

However, when we examine the matrix generalization of the moments (48), it becomes obvious that only a *diagonal* version of $\underset{\sim}{v}_0$ is acceptable; the only way to achieve this is to define a *new* matrix version of $\underset{\sim}{y}_0$:

$$\underset{\sim}{Y}_0 = \mathrm{diag}\ (\underset{\sim}{y}_0) = \begin{pmatrix} y_{10} & & & 0 \\ & y_{20} & & \\ & & \ddots & \\ 0 & & & y_{p0} \end{pmatrix} \ ; \quad \underset{\sim}{y}_0 = \underset{\sim}{Y}_0 \cdot \underset{\sim}{1} \ ; \tag{61}$$

and to take as definition of $\underset{\sim}{V}_0$ any of the equivalent forms:

$$\underset{\sim}{V}_0 = \underset{\sim}{R}_0 - \underset{\sim}{N}_0^{-1}\underset{\sim}{Y}_0\underset{\sim}{Y}_0' = \underset{\sim}{R}_0 - \underset{\sim}{Y}_0\underset{\sim}{N}_0^{-1}\underset{\sim}{Y}_0' = \underset{\sim}{R}_0 - \underset{\sim}{Y}_0\underset{\sim}{Y}_0'\underset{\sim}{N}_0^{-1} \ , \tag{62}$$

then replacing $\underset{\sim}{\pi}'\underset{\sim}{v}_0$ by $\mathrm{tr}\ (\underset{\sim}{\Pi}\underset{\sim}{V}_0)$.

After transformation, we have:

$$p(\underset{\sim}{\Omega}) \propto \prod_{i=1}^{p} [(\underset{\sim}{A}'\underset{\sim}{\Omega}\underset{\sim}{A})_i]^{\frac{1}{2}(n_{i0}+1)} \exp\left\{-\frac{1}{2}\ \mathrm{tr}\ (\underset{\sim}{\Omega}\underset{\sim}{U}_0)\right\} \ , \tag{63}$$

which might be called a linearly-dependent Wishart density. The transformed precision-parameter $\underset{\sim}{U}_0$ is given by:

$$\underset{\sim}{U}_0 = \underset{\sim}{Q}_0 - \underset{\sim}{N}^{-1}\underset{\sim}{X}_0\underset{\sim}{X}_0' = \underset{\sim}{Q}_0 - \underset{\sim}{X}_0\underset{\sim}{N}_0^{-1}\underset{\sim}{X}_0' = \underset{\sim}{Q}_0 - \underset{\sim}{X}_0\underset{\sim}{X}_0'(\underset{\sim}{N}^{-1})' \ , \tag{64}$$

(the $\underset{\sim}{N}_0$ in the middle form is *not* a typographical error), where we define a new matrix of hyperparameters:

$$\underset{\sim}{X}_0 = \underset{\sim}{A}\underset{\sim}{Y}_0 = \underset{\sim}{A}\ \mathrm{diag}\left(\underset{\sim}{A}^{-1}\underset{\sim}{x}_0\right) \quad ; \quad \underset{\sim}{X}_0 \cdot \underset{\sim}{1} = \underset{\sim}{x}_0 \ ; \tag{65}$$

which, although neither diagonal nor symmetric, has sums over columns equal to $\underset{\sim}{x}_0$.

Now, we can safely rewrite the means in (48), and obtain, finally:

$$E\{\tilde{\underset{\sim}{\Omega}}\} = (\underset{\sim}{N}' + 3\underset{\sim}{I})\underset{\sim}{U}_0^{-1} \quad ; \quad E\{\tilde{\underset{\sim}{\Omega}}^{-1}\} = \underset{\sim}{E} = (\underset{\sim}{N} + \underset{\sim}{I})^{-1}\underset{\sim}{U}_0 \ . \tag{66}$$

Explicit forms for the covariances of the covariance can be found from (48), but give formulae involving $\underset{\sim}{A}$.

It is also difficult to use (46) to produce a simple formula for $p(\underset{\sim}{\mu})$, similar to (37), which might be termed a linearly-dependent Student-t; generalizations of this and more complicated type often arise in multivariate analysis. Similar remarks apply to the exact form for our outcome density, $p(\underset{\sim}{x})$.

However, moments corresponding to (49) (50) follow directly by transformation:

$$E\{\tilde{\underset{\sim}{\mu}}\} = \underset{\sim}{N}^{-1}\underset{\sim}{x}_0 \quad ; \quad \underset{\sim}{V}\{\tilde{\underset{\sim}{\mu}}\} = \underset{\sim}{D} = \underset{\sim}{N}^{-1}(\underset{\sim}{N} + \underset{\sim}{I})^{-1}\underset{\sim}{U}_0 \ ; \tag{67}$$

and

$$E\{\tilde{\underset{\sim}{x}}\} = \underset{\sim}{N}^{-1}\underset{\sim}{x}_0 \quad ; \quad \underset{\sim}{V}\{\tilde{\underset{\sim}{x}}\} = \underset{\sim}{N}^{-1}\underset{\sim}{U}_0 \ . \tag{68}$$

These are exactly equivalent to the Ando-Kaufman results, with the exception that $\underset{\sim}{N}$ is there replaced by $n_{00}\underset{\sim}{I}$, and $\underset{\sim}{U}_0$ given by (36), instead of (64). The variance-covariance matrices are symmetric, since (cf. (56)), $\underset{\sim}{N}\underset{\sim}{U}_0 = \underset{\sim}{U}_0\underset{\sim}{N}'$, and similarly for other powers of $\underset{\sim}{N}$ or $\underset{\sim}{N}^{-1}$.

8. LDMEF CREDIBILITY FORECASTS

We are now ready to analyze the updating to the LDMEF prior, and then to obtain the forecasts from (68). Starting from $\mathcal{D}_y = \{\underset{\sim}{y}_t \; ; \; t = 1,2, \ldots, n\}$, we see that we again get quadratic difficulties (similar to the definition of $\underset{\sim}{v}_0$), since $r_{i0} \leftarrow r_{i0} + \sum_{t=1}^{n} y_{it}^2$. This leads us to define expanded versions of $\underset{\sim}{y}_t$ and $\underset{\sim}{x}_t$, corresponding to (61) (65):

$$\begin{gathered} \underset{\sim}{Y}_t = \operatorname{diag}(\underset{\sim}{y}_t) \quad ; \quad \underset{\sim}{y}_t = \underset{\sim}{Y}_t \underset{\sim}{1} \; ; \\ \underset{\sim}{X}_t = \underset{\sim}{A}\underset{\sim}{Y}_t = \underset{\sim}{A} \operatorname{diag}\left(\underset{\sim}{A}^{-1}\underset{\sim}{x}_t\right) \quad ; \quad \underset{\sim}{x}_t = \underset{\sim}{X}_t \cdot \underset{\sim}{1} \qquad (t = 1,2, \ldots, n) \; . \end{gathered} \tag{69}$$

It follows easily that, posterior-to-data $\mathcal{D}_x$, the LDMEF priors (55) (59) and the various moments merely require the updating:

$$\begin{gathered} \underset{\sim}{N} \leftarrow \underset{\sim}{N} + n\underset{\sim}{I} \; : \; \underset{\sim}{x}_0 \leftarrow \underset{\sim}{x}_0 + \sum_{t=1}^{n} \underset{\sim}{x}_t \quad ; \quad \underset{\sim}{X}_0 \leftarrow \underset{\sim}{X}_0 + \sum_{t=1}^{n} \underset{\sim}{X}_t \; ; \\ \underset{\sim}{Q}_0 \leftarrow \underset{\sim}{Q}_0 + \sum_{t=1}^{n} \underset{\sim}{X}_t \underset{\sim}{X}_t' \; ; \end{gathered} \tag{70}$$

which shall be compared with (38) and (29). While we have succeeded in introducing the enriched time constants, we see that we require a new, expanded form for updating $\underset{\sim}{Q}_0$, and hence $\underset{\sim}{U}_0$.

The credible mean is unaffected by this. Setting $\underset{\sim}{m} = \underset{\sim}{N}^{-1}\underset{\sim}{x}_0$ to coincide with (8), we find that the first formula in (68) gives the predictor:

$$E\{\tilde{\underset{\sim}{x}} \mid \mathcal{D}_x\} = (\underset{\sim}{I} - \underset{\sim}{Z})\underset{\sim}{m} + \underset{\sim}{Z}\bar{\underset{\sim}{x}} \; , \tag{71}$$

as expected.

But to update the covariance matrix, we find that we also need expanded versions of $\underset{\sim}{m}$ and $\bar{\underset{\sim}{x}}$, as follows:

$$\underset{\sim}{M} = \underset{\sim}{N}^{-1}\underset{\sim}{X}_0 = \underset{\sim}{A}\ \mathrm{diag}\ (\underset{\sim}{A}^{-1}\underset{\sim}{m}) \quad ; \quad \underset{\sim}{M}\underset{\sim}{1} = \underset{\sim}{m}\ ;$$
$$\bar{\underset{\sim}{X}} = \frac{1}{n}\sum_{t=1}^{n} \underset{\sim}{X}_t = \underset{\sim}{A}\ \mathrm{diag}\ (\underset{\sim}{A}^{-1}\bar{\underset{\sim}{x}}) \quad ; \quad \bar{\underset{\sim}{X}}\underset{\sim}{1} = \bar{\underset{\sim}{x}}\ . \tag{72}$$

After some algebra, and liberal use of symmetry, we obtain:

$$\underset{\sim}{V}\{\tilde{\underset{\sim}{x}} \mid \mathcal{D}_x\} = (\underset{\sim}{I} - \underset{\sim}{Z})\underset{\sim}{C} + \underset{\sim}{Z}\underset{\sim}{T}(\mathcal{D}_x) + \underset{\sim}{Z}(\underset{\sim}{I} - \underset{\sim}{Z})(\bar{\underset{\sim}{X}} - \underset{\sim}{M})(\bar{\underset{\sim}{X}} - \underset{\sim}{M})'\ , \tag{73}$$

with

$$\underset{\sim}{T}(\mathcal{D}_x) = \frac{1}{n}\sum_{t=1}^{n} (\underset{\sim}{X}_t - \bar{\underset{\sim}{X}})(\underset{\sim}{X}_t - \bar{\underset{\sim}{X}})'\ , \tag{74}$$

which should be compared to the result (40) with the Ando-Kaufman prior. We see that, as a price for working with a full credibility matrix $\underset{\sim}{Z}$, we must use the expanded forms for the inter-risk and intra-risk statistics. (73) was given incorrectly in Jewell (1974b) because of the dimensionality problems discussed earlier.

This explains why $\underset{\sim}{T}(\mathcal{D}_x)$ is the correct sufficient statistic, as $n \to \infty$, for the dispersion or precision of $\tilde{\underset{\sim}{x}}$, instead of the more usual sufficient statistic, $\underset{\sim}{S}(\mathcal{D}_x) = n^{-1}\sum(\underset{\sim}{x}_t - \bar{\underset{\sim}{x}})(\underset{\sim}{x}_t - \bar{\underset{\sim}{x}})'$. The latter is sufficient for $\tilde{\underset{\sim}{\Omega}}$ only if $\tilde{\underset{\sim}{\Omega}}$ is of full rank, while in our case, it was constructed from a lower-dimensional vector, $\tilde{\underset{\sim}{\pi}}$.

We give two additional formulae not in the original paper:

$$E\underset{\sim}{V}\{\tilde{\underset{\sim}{x}} \mid \tilde{\underset{\sim}{\Omega}} \mid \mathcal{D}_x\} = E\left\{\tilde{\underset{\sim}{\Omega}}^{-1} \mid \mathcal{D}_x\right\} = (\underset{\sim}{I} - \underset{\sim}{Z}^{+})\underset{\sim}{E} + \underset{\sim}{Z}^{+}\underset{\sim}{T}(\mathcal{D}_x) + \underset{\sim}{Z}^{+}(\underset{\sim}{I} - \underset{\sim}{Z})(\underset{\sim}{X} - \underset{\sim}{M})(\underset{\sim}{X} - \underset{\sim}{M})' \; ; \tag{75}$$

$$\underset{\sim}{V}E\{\tilde{\underset{\sim}{x}} \mid \tilde{\underset{\sim}{\mu}} \mid \mathcal{D}_x\} = \underset{\sim}{V}\{\tilde{\underset{\sim}{\mu}} \mid \mathcal{D}_x\} = (\underset{\sim}{I} - \underset{\sim}{Z})\underset{\sim}{D} - \frac{1}{n}\underset{\sim}{Z}\underset{\sim}{Z}^{+}[-\underset{\sim}{E} + \underset{\sim}{T}(\mathcal{D}_x) + (\underset{\sim}{I} - \underset{\sim}{Z})(\underset{\sim}{X} - \underset{\sim}{M})(\underset{\sim}{X} - \underset{\sim}{M})'] \; ; \tag{76}$$

where $\underset{\sim}{Z}^{+}$ is a credibility matrix with larger time constants:

$$\underset{\sim}{Z}^{+} = n(\underset{\sim}{N} + \underset{\sim}{I} + n\underset{\sim}{I})^{-1} \; . \tag{77}$$

(75) is the persistent part of (73), while the uncertainty about $\tilde{\underset{\sim}{\mu}}$ vanishes almost surely (in the non-degenerate case) as $n \to \infty$.

9. RANDOM PRECISION ONLY

Our new results can be most easily seen in isolation, by considering the case where only $\tilde{\Omega}$ is random, and we set $\underset{\sim}{m} \equiv 0$ for convenience. The problem of predicting $\tilde{\Omega}^{-1}$ is of continuing interest in the Bayesian literature, see, e.g., Dickey, Lindley, and Press (1981).

For the thin natural-conjugate prior, we obtain easily:

$$p(\underset{\sim}{\Omega}) \propto |\underset{\sim}{\Omega}|^{\frac{1}{2}n_{00}} \exp\left\{-\frac{1}{2}\operatorname{tr}(\underset{\sim}{\Omega}\underset{\sim}{U}_0)\right\} , \tag{78}$$

that is, $\tilde{\Omega} \sim \mathcal{W}_p(n_{00} + p + 1 ; \underset{\sim}{U}_0)$ with $\frac{1}{2}(p^2 + p + 2)$ hyperparameters $(n_{00}, \underset{\sim}{U}_0)$ to be estimated ($\underset{\sim}{U}_0$ must be symmetric, not for the trace, but for the moments below). This may be compared with (35). Updating is as in the first and third formulae of (38), so that the thin forecast of the covariance of the observations, corresponding to (40), is simply:

$$\underset{\sim}{V}\{\tilde{\underset{\sim}{x}} \mid \mathcal{D}_x\} = E\left\{\tilde{\underset{\sim}{\Omega}}^{-1} \mid \mathcal{D}_x\right\} = (1 - z_{00})\underset{\sim}{C} + z_{00}\left[\frac{1}{n}\sum_{t=1}^{n} \underset{\sim}{x}_t\underset{\sim}{x}_t'\right] , \tag{79}$$

where $\underset{\sim}{C} = n_{00}^{-1}\underset{\sim}{U}_0$. This result is well-known, although not usually expressed in shrinkage form (see, e.g., Press (1981)). (Remember $\underset{\sim}{C} = \underset{\sim}{E}$, and $\underset{\sim}{D}$, $\underset{\sim}{m}$ vanish).

For the enriched prior, we follow the appropriate steps of Section 6 and find that:

$$p(\underset{\sim}{\Omega}) \propto \prod_{i=1}^{p} [(\underset{\sim}{A}'\underset{\sim}{\Omega}\underset{\sim}{A})_i]^{\frac{1}{2}n_{i0}} \exp\left\{-\frac{1}{2}\operatorname{tr}(\underset{\sim}{\Omega}\underset{\sim}{U}_0)\right\} \tag{80}$$

for some $\tilde{\Omega}$ constrained by (44) or (56) or (57), that is, of rank p , and with $\underset{\sim}{U}_0 = \underset{\sim}{Q}_0 = \underset{\sim}{A}\underset{\sim}{R}_0\underset{\sim}{A}'$, where $\underset{\sim}{R}_0$ is diagonal, so that (56) applies. We

thus have the 2p initial hyperparameters $\{\underset{\sim}{n}_0;\underset{\sim}{r}_0\}$ for $\tilde{\underset{\sim}{\pi}}$, plus the effective terms in the transformation matrix $\underset{\sim}{A}$ (see below).

After arguments similar to those of Section 7, we find the enriched covariance forecast to be:

$$\underset{\sim}{V}\{\tilde{\underset{\sim}{x}} \mid \mathcal{D}_x\} = (\underset{\sim}{I} - \underset{\sim}{Z})\underset{\sim}{C} + \underset{\sim}{Z}\,\frac{1}{n}\left[\sum_{t=1}^{n} \underset{\sim}{X}_t\underset{\sim}{X}_t'\right], \tag{81}$$

which should be compared with (73). $\underset{\sim}{C}$ is now $\underset{\sim}{N}^{-1}\underset{\sim}{U}_0$, and again the enriched prior makes the credibility factor matrix full-dimensional, but changes the sufficient statistics to their expanded versions, using $\underset{\sim}{X}_t\underset{\sim}{X}_t'$ instead of $\underset{\sim}{x}_t\underset{\sim}{x}_t'$.

One can determine second moments of $\tilde{\underset{\sim}{\Omega}}$ or $\underset{\sim}{\Omega}^{-1} = \underset{\sim}{\Sigma}$ through transformations of the formulae in the Appendix.

10. PROPERTIES OF THE TRANSFORMATION MATRIX

Having produced an enriched prior and the associated prediction formulae, we now turn to the problem of specifying the needed hyperparameters. But first, we must consider how to find the appropriate transformation matrix $\underset{\sim}{A}$.

In some situations, there may be a natural choice for $\underset{\sim}{A}$. For example, in their 1981 paper, Dickey, Lindley, and Press assume that (in our notation) $\underset{\sim}{C}$ in (79) is of interclass form, $\underset{\sim}{C} = \sigma^2[(1 - p)\underset{\sim}{I} + p\underset{\sim}{1}\underset{\sim}{1}']$, and note that there exists an orthogonal matrix $\underset{\sim}{\Gamma}$ which diagonalizes $\underset{\sim}{C}$ and $\underset{\sim}{U}_0$ via $\underset{\sim}{\Gamma}\underset{\sim}{C}\underset{\sim}{\Gamma}' = \text{diag}\ (\alpha, \beta, \ldots, \beta)$, $\alpha = \sigma^2 + (p - 1)\beta$, $\beta = \sigma^2(1 - p)$. From (54), we see that this essentially defines $\underset{\sim}{A} = \underset{\sim}{\Gamma}^{-1} = \underset{\sim}{\Gamma}'$ as a particular orthogonal matrix, and this shows how to calculate the expanded observations $\underset{\sim}{X}_t$ in (81), and the possible forms for $\underset{\sim}{N}$, which here becomes symmetric. In fact, for this interclass $\underset{\sim}{C}$, there are many possible $\underset{\sim}{A}$, as the first column of $\underset{\sim}{A}$ is $p^{-\frac{1}{2}}\underset{\sim}{1}$, and the other columns are any set of $(p - 1)$ mutually orthogonal, normed vectors, also orthogonal to $\underset{\sim}{1}$; this gives $\frac{1}{2}(p - 1)(p - 2)$ free choices.

In the more general situation, we may wish to test a given $\underset{\sim}{N}$, to see if it is of permitted form. Letting $\{\underset{\sim}{a}_j\}$ be the column vectors of $\underset{\sim}{A}$, we see that (54) can be rewritten:

$$\underset{\sim}{N}\underset{\sim}{a}_j = n_{j0}\underset{\sim}{a}_j , \tag{82}$$

which means that:

(1) the hyperparameters $\{n_{i0}\}$ of the independent priors (45) are the *eigenvalues* $\{\nu_i\}$ of $\underset{\sim}{N}$, referred to in Section 2; these must be positive (or, greater than unity) if we want the moments (48) to be well-defined;

(2) the columns of $\underset{\sim}{A}$ are the (right) *eigenvectors* of $\underset{\sim}{N}$, which must be mutually independent, since we have assumed $\underset{\sim}{A}$ is of rank p; however, they need not be mutually orthogonal, since $\underset{\sim}{N}$ is not necessarily symmetric;

(3) from the fact that (54) (82) are similarity transforms, it follows that:

$$|\underset{\sim}{N}| = |\underset{\sim}{N}_0| = \prod_{i=1}^{p} n_{i0} > 0 \quad (\text{or}, \ > 1) \ ;$$

$$\text{tr}\ (\underset{\sim}{N}) = \text{tr}\ (\underset{\sim}{N}_0) = \sum_{i=1}^{p} n_{ii} = \sum_{i=1}^{p} n_{i0} > 0 \quad (\text{or}, \ > p) \ . \tag{83}$$

This interpretation shows also that we can *norm* the column vectors of the transformation, by using a matrix $\underset{\sim}{A}^*$, with columns $\underset{\sim}{a}_j^* = k_j \underset{\sim}{a}_j$, (which is equivalent to a *scale change* of k_i in the underlying y_i, used with the original $\underset{\sim}{A}$). The matrix of time constants, $\underset{\sim}{N}$, and the process for forming $\underset{\sim}{X}_t$, $\underset{\sim}{X}$, and $\underset{\sim}{M}$ are unaffected by this change, while the hyperparameters (x_0, Q_0) merely reflect the underlying scale changes. Since our prediction formulae are in terms of the final moments in $\mathcal{X}$, $(\underset{\sim}{m}, \underset{\sim}{D}, \underset{\sim}{E}, \underset{\sim}{C})$, this scale change does not affect (71) (73) or (81), either. In other words, the use of an arbitrary, invertible $\underset{\sim}{A}$ can only introduce $p^2 - p$ effective new hyperparameters into the enriched credibility formulae; the remaining p column norms merely make "nuisance" scale changes.

An important special case occurs if $\underset{\sim}{A}$ is an *orthogonal* matrix, so that $\underset{\sim}{A}^{-1} = \underset{\sim}{A}'$, and all the transformations from $\underset{\sim}{N}$, $\underset{\sim}{\Omega}$, and $\underset{\sim}{Q}_0$ to and from their diagonal counterparts become the same *orthogonal transformation*; this makes $\underset{\sim}{N}$ symmetric, and from (56) (57), $\underset{\sim}{N}$, $\underset{\sim}{\Omega}$, and $\underset{\sim}{Q}_0$ all commute with

each other. The converse is usually only partially true, since all "normal" real matrices $(\underset{\sim}{N}'\underset{\sim}{N} = \underset{\sim}{N}\underset{\sim}{N}')$ also have orthogonal reductions. But, in our case, we require the $\{n_{i0}\}$ to be real, and this means $\underset{\sim}{N}$ must be symmetric (see Bellman (1960) or Nobel (1969)). Therefore, $(\underset{\sim}{A}^{-1} = \underset{\sim}{A}')$ iff $(\underset{\sim}{N} = \underset{\sim}{N}')$ for our problem.

The question of whether or not $\underset{\sim}{N}$ is symmetric is tied up with the reductions of $\underset{\sim}{E}$ and $\underset{\sim}{D}$, which, being symmetric, always have orthogonal diagonalizations. If the orthogonal transformation matrix is the *same* for both $\underset{\sim}{E}$ and $\underset{\sim}{D}$, then they are said to be *simultaneously* orthogonally diagonalized (Bellman (1960)) which can only occur if they commute $(\underset{\sim}{E}\underset{\sim}{D} = \underset{\sim}{D}\underset{\sim}{E})$, i.e., if $\underset{\sim}{N} = \underset{\sim}{N}'$.

11. SPECIFICATION OF HYPERPARAMETERS

Eventually, every Bayesian analysis must address the problem of specification of the prior hyperparameters. Several authors have mentioned the difficulties of forming a prior opinion about the Wishart prior (35), especially because of its thinness. (63) is a little easier to visualize, as it is formed by a linear transformation from independent Gamma densities (45), yet here we have the problem of reduced rank, and the required inter-relations (56) (57) between $\tilde{\underset{\sim}{\Omega}}$ and the hyperparameters.

The point of view we wish to emphasize in this section is that, if the objective of the Bayesian analysis is to forecast the mean and variance of $\tilde{\underset{\sim}{x}}$, we can focus our attention on the estimation of parameters which are pretty much in the space X, and which have natural classical estimators when large amounts of collateral data are available. (However, we will not discuss the problem of actually making these estimates: see, inter alia, De Vylder (1978) (1981), Norberg (1980) (1982), Zehnwirth (1981) and Sundt (1981)). Counting the number of hyperparameters to be specified also indicates the additional modelling flexibility of an enriched prior.

As a warm-up, let us first examine the basic credibility model of Section 2, where $\tilde{\underset{\sim}{\mu}}$ is random, but $\underset{\sim}{V}\{\tilde{\underset{\sim}{x}} \mid \tilde{\underset{\sim}{\mu}}\} = \underset{\sim}{W}^{-1} = \underset{\sim}{E}$ is given. In the simple natural conjugate prior (19), we see that there are $p + 1$ hyperparameters $(n_{00};\underset{\sim}{m})$ to be specified; but $\underset{\sim}{m} = E\{\tilde{\underset{\sim}{x}}\}$, and n_{00} is got from $n_{00}\underset{\sim}{I} = \underset{\sim}{D}^{-1}\underset{\sim}{W}^{-1}$, where $\underset{\sim}{D} = \underset{\sim}{V}\{\underset{\sim}{\mu}\}$ must be similar to $\underset{\sim}{E}$, so these parameters are easily visualized. The enriched version (30) has $p + p^2$ hyperparameters $(\underset{\sim}{m},\underset{\sim}{N})$, but these are also easily visualized, as again $\underset{\sim}{m} = E\{\tilde{\underset{\sim}{x}}\}$, but $\underset{\sim}{N}$ is got from $\underset{\sim}{N} = \underset{\sim}{D}^{-1}\underset{\sim}{W}^{-1}$, with $\underset{\sim}{D} = \underset{\sim}{V}\{\tilde{\underset{\sim}{\mu}}\}$ now an arbitrary, symmetric positive definite matrix to be estimated. In other words, apart from $\underset{\sim}{m}$ and $\underset{\sim}{C} = \underset{\sim}{V}(\tilde{\underset{\sim}{x}})$

the only additional prior specification problem involves the split of variance into its inter-risk and intra-risk components, one of which is given. The modelling gain in using the enriched version is $p^2 - 1$ additional parameters.

Passing now to the case when both $\underset{\sim}{\mu}$, $\underset{\sim}{\Omega}$ are the $\frac{1}{2}(p^2 + 3p)$ random parameters, we recall that the Ando-Kaufman prior (33) requires $\frac{1}{2}(p^2 + 3p + 2)$ hyperparameters $\{n_{00}, \underset{\sim}{x}_0, \underset{\sim}{Q}_0\}$. The mean and the variance of (37) give $\frac{1}{2}(p^2 + 3p)$ of these: $E\{\tilde{\underset{\sim}{\mu}}\} = \underset{\sim}{m} = n_{00}^{-1}\underset{\sim}{x}_0$, and $\underset{\sim}{V}(\tilde{\underset{\sim}{\mu}}) = \underset{\sim}{D} = n_{00}^{-1}(n_{00} + 1)^{-1}\underset{\sim}{U}_0$, leaving only *one* degree of freedom, since $E\underset{\sim}{V}\{\tilde{\underset{\sim}{x}} \mid \underset{\sim}{\Omega}\} = \underset{\sim}{E} = n_{00}\underset{\sim}{D}$! Alternately, we could ignore $\underset{\sim}{D}$, fix $\frac{1}{2}(p^2 + p)$ coefficients from the mean of $\tilde{\underset{\sim}{\Omega}}$ or $\tilde{\underset{\sim}{\Omega}}^{-1} = \underset{\sim}{\Sigma}$ and then estimate *one* second moment of $\tilde{\underset{\sim}{\Omega}}$ or $\tilde{\underset{\sim}{\Sigma}}$ from the formulae in the Appendix; all other second moments are then determined! In short, there is one more hyperparameter than the number of parameters, no matter the size of p.

Our enriched prior (55) can be evaluated at several different stages. In terms of $\underset{\sim}{y}$ and $(\underset{\sim}{\lambda}, \underset{\sim}{\pi})$, we see from (45) that, for $2p$ random parameters, we have first $3p$ hyperparameters (n_0, y_0, r_0), augmented by the effective number of elements in $\underset{\sim}{A}$ ($p^2 - p$, by a previous argument), for a total of $p^2 + 2p$ hyperparameters. In terms of (55), we count first the $p + p^2 + \frac{1}{2}p(p + 1)$ hyperparameters $\{\underset{\sim}{N}, \underset{\sim}{x}_0, \underset{\sim}{Q}_0\}$, which must then be reduced by the $\frac{1}{2}(p^2 - p)$ constraints (56), for a total of $p^2 + 2p$ effective hyperparameters. Finally, in the specification the author prefers, we think of the problem of specifying prior estimates of $\underset{\sim}{m}$, $\underset{\sim}{E}$, and $\underset{\sim}{D}$. We then compute $\underset{\sim}{N} = \underset{\sim}{E}\underset{\sim}{D}^{-1}$, $\underset{\sim}{x}_0 = \underset{\sim}{N}\underset{\sim}{m}$, and $\underset{\sim}{U}_0 = (\underset{\sim}{N} + \underset{\sim}{I})\underset{\sim}{E}$. The eigenvalues of $\underset{\sim}{N}$ are then computed, and if these are negative (or less than unity), then the assumptions about $\underset{\sim}{E}$, $\underset{\sim}{D}$ must be inconsistent with the LDMEF model. Usually, however, there will be no difficulty at this point, and $\underset{\sim}{A}$ is

determined as the matrix of eigenvectors of $\underset{\sim}{N}$, (which will be orthogonal if $\underset{\sim}{N}$ is symmetric). The determinations of $\underset{\sim}{X}_t$ $(t = 0,1, \ldots, n)$, $\bar{\underset{\sim}{X}}$, and $\underset{\sim}{M}$ are immediate from (69) (72), whence one can find $\underset{\sim}{Q}_0$ from (64), and the credibility forecasts from (71) (73). ($\underset{\sim}{m}$, $\underset{\sim}{E}$, and $\underset{\sim}{D}$) give also $p + \frac{1}{2}(p^2 + p) + \frac{1}{2}(p^2 + p) = p^2 + 2p$ degrees of freedom, p^2 more than the number of random parameters. By comparison with the Ando-Kaufman prior, we see that $\frac{1}{2}(p^2 + p - 2)$ more hyperparameters have been introduced.

Finally, if only the precision $\tilde{\underset{\sim}{\Omega}}$ $\left(\text{or covariance } \tilde{\underset{\sim}{\Sigma}}\right)$ is random, (78) shows that, for the $\frac{1}{2}(p^2 + p)$ random parameters, the Wishart prior requires specification of $\frac{1}{2}(p^2 + p + 2)$ hyperparameters $(n_{00}, \underset{\sim}{Q}_0)$, i.e., $\underset{\sim}{C}$ and *one* additional component of say, the variance of $\tilde{\underset{\sim}{\Sigma}}$. The enriched prior, on the other hand, only contains p independent random variables, but utilizes $p^2 + p$ hyperparameters, computed in the various ways indicated above. Since $\underset{\sim}{m} = \underset{\sim}{D} = 0$, this suggests that the analyst could estimate $\underset{\sim}{C} = \underset{\sim}{E} = \underset{\sim}{N}^{-1}\underset{\sim}{U}_0$, but must obtain the remaining $\frac{1}{2}(p^2 + p)$ parameters from the third moments of $\tilde{\underset{\sim}{x}}$, the second moments of $\tilde{\underset{\sim}{\Sigma}}$, or else some physical interpretation of $\underset{\sim}{N}$ or $\underset{\sim}{A}$.

Example:

To illustrate the issues raised in the previous sections, we consider a two-dimensional example where $\underset{\sim}{m}$, and the two components of the covariance have been estimated:

$$\underset{\sim}{D} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{5}{6} \end{bmatrix} \quad ; \quad \underset{\sim}{E} = \begin{bmatrix} e_{11} & e_{12} \\ e_{12} & e_{22} \end{bmatrix} .$$

For $\underset{\sim}{E}$ to be positive definite, e_{11} , e_{12} , and $\left(e_{11}e_{22} - e_{12}^2\right)$ must be positive. Then, we find easily that $\underset{\sim}{D}$ is diagonalized by $\underset{\sim}{\Gamma}_D$, with eigenvalues $\underset{\sim}{\delta}$, where:

$$\underset{\sim}{\Gamma}_D'\underset{\sim}{D}\underset{\sim}{\Gamma}_D = \text{diag}\ (\underset{\sim}{\delta}) \quad ; \quad \underset{\sim}{\Gamma}_D = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 & +1 \\ -1 & 2 \end{bmatrix} \quad ; \quad \underset{\sim}{\delta} = \left[\frac{1}{6}, 1\right]' \ ;$$

and $\underset{\sim}{E}$ is diagonalized by $\underset{\sim}{\Gamma}_E$, with eigenvalues $\underset{\sim}{\varepsilon}$, where:

$$\underset{\sim}{\Gamma}_E'\underset{\sim}{E}\underset{\sim}{\Gamma}_E = \text{diag}\ (\underset{\sim}{\varepsilon}) \quad ; \quad \underset{\sim}{\Gamma}_E = \begin{bmatrix} k_1 e_{12} & k_2(\varepsilon_2 - e_{22}) \\ k_1(\varepsilon_1 - e_{11}) & k_2 e_{12} \end{bmatrix} ;$$

$$\varepsilon_{1,2} = \frac{1}{2}\ (e_{11} + e_{22}) \pm \sqrt{e_{12}^2 + \frac{1}{4}\ (e_{11} - e_{22})^2} \quad ; \quad k_i = \left[e_{12}^2 + (\varepsilon_i - e_{ii})^2\right]^{-\frac{1}{2}} \qquad (i = 1,2) \ .$$

This makes the matrix of time constants, $\underset{\sim}{N}$, and the credibility matrix, $\underset{\sim}{Z}$,

$$\underset{\sim}{N} = \begin{bmatrix} 5e_{11} - 2e_{12} & 2e_{12} - 2e_{11} \\ 5e_{12} - 2e_{22} & 2e_{22} - 2e_{12} \end{bmatrix} \quad ; \quad \underset{\sim}{Z} = \frac{1}{\Delta(n)} \begin{bmatrix} n(n + 2e_{22} - 2e_{12}) & -2n(e_{12} - e_{11}) \\ -n(5e_{12} - 2e_{22}) & n(n + 5e_{11} - 2e_{12}) \end{bmatrix} ;$$

$$\Delta(n) = n^2 + n[5e_{11} + 2e_{22} - 4e_{12}] + 6\left[e_{11}e_{22} - e_{12}^2\right] .$$

The eigenvalues of $\underset{\sim}{N}$, $\underset{\sim}{\nu} = [\nu_1,\nu_2]'$, are the two roots of $\Delta(-\nu) = 0$; in order that they both be real and positive:

$$5e_{11} + 2e_{22} - 4e_{12} > 0 \quad ; \quad e_{11}e_{12} > e_{12}^2 \ .$$

In this case, the second condition dominates, and is automatically satisfied if $\underset{\sim}{E} > 0$. From the eigenvectors, we can form

$$\underset{\sim}{A} = \begin{bmatrix} 2e_{12} - 2e_{11} & (\nu_2 + 2e_{12} - 2e_{22}) \\ \nu_1 + 2e_{12} - 5e_{11} & 5e_{12} - 2e_{22} \end{bmatrix} ,$$

or any similar $\underset{\sim}{A}$ with normed columns, so that $\underset{\sim}{A}^{-1}\underset{\sim}{N}\underset{\sim}{A} = \text{diag}\ (\underset{\sim}{\nu}) = \underset{\sim}{N}_0$.

If we use the Ando-Kaufman prior (33), we can set $\underset{\sim}{m}$ and $\underset{\sim}{D}$ as above, but we require that $\underset{\sim}{E} = n_{00}\underset{\sim}{D}$, which means that:

$$e_{11} = e_{12} \quad ; \quad e_{22} = \frac{5}{2}\, e_{12} \quad ; \quad n_{00} = 3e_{12} \ ;$$

and that $e_{12} > 0$ or 1. From these we get the 6 hyperparameters $(n_{00},\underset{\sim}{x}_0,\underset{\sim}{Q}_0)$ for the prior; the 3 precision-parameters, $(\tilde{\omega}_{11},\tilde{\omega}_{12},\tilde{\omega}_{22})$, are constrained by the positive-definiteness requirement $\tilde{\omega}_{11}\tilde{\omega}_{22} > \tilde{\omega}_{12}^2$. From the above we see that $\underset{\sim}{\varepsilon} = \left[\frac{1}{2}\, e_{12};3e_{12}\right]'$, and $\underset{\sim}{\Gamma}_E = \underset{\sim}{\Gamma}_D$ (after possibly permuting columns), and $\nu_i = \varepsilon_i/\delta_i = 3e_{12}$ $(i = 1,2)$. If we use the prior (55) with the above values of (e_{ij}) , then $\underset{\sim}{N}$ is already diagonalized, and *any* invertible $\underset{\sim}{A}$ works, so we might as well use $\underset{\sim}{A} = \underset{\sim}{I}$. Note also that the (single) constraint (44) or (57) is vacuous, in this case. A typical term in the variance forecasts (73) and (81) would use:

$$\underset{\sim}{x}_t\underset{\sim}{x}_t' = \begin{bmatrix} x_{1t}^2 & x_{1t}x_{2t} \\ x_{1t}x_{2t} & x_{2t}^2 \end{bmatrix} \ ; \ \underset{\sim}{X}_t\underset{\sim}{X}_t' = \begin{bmatrix} x_{1t}^2 & 0 \\ 0 & x_{2t}^2 \end{bmatrix} ;$$

showing the underlying independence in this case, since $\underset{\sim}{Z}$ is diagonal.

All other choices of the (e_{ij}) require the more general prior (55). An important special case occurs when:

$$e_{22} = e_{11} + \frac{3}{2} e_{12} \qquad (e_{11} \neq e_{12}) ,$$

with $e_{11} > \frac{1}{2} e_{12}$ to make $\underset{\sim}{E} > 0$. Then $\underset{\sim}{\varepsilon} = \left[e_{11} - \frac{1}{2} e_{12} ; e_{11} + 2e_{12}\right]'$ and we again find $\underset{\sim}{\Gamma}_E = \underset{\sim}{\Gamma}_D$. This means that $\underset{\sim}{N}$ is symmetric, and we find the underlying time constants to be:

$$\underset{\sim}{\nu} = [6e_{11} - 3e_{12} ; e_{11} + 2e_{12}]' \quad ; \quad (\nu_i = \varepsilon_i/\delta_i \; ; \; i = 1,2) \; ;$$

both now greater than zero if $e_{11} > -2e_{12}$. In fact, $\underset{\sim}{A}$ can be normalized so that $\underset{\sim}{A} = \underset{\sim}{\Gamma}_E = \underset{\sim}{\Gamma}_D = (\underset{\sim}{A}^{-1})'$! A typical term in the variance forecast (81) would be:

$$\underset{\sim}{X}_t \underset{\sim}{X}_t' = \frac{1}{25} \left\{ x_{1t}^2 \begin{bmatrix} 5 & -6 \\ -6 & 8 \end{bmatrix} + x_{1t} x_{2t} \begin{bmatrix} -4 & 14 \\ 14 & 12 \end{bmatrix} + x_{2t}^2 \begin{bmatrix} 5 & 7 \\ 7 & 17 \end{bmatrix} \right\} .$$

All other cases give more general results. For example, if

$$e_{11} = e_{12} = Ke_{12} ,$$

then $K > 1$ to make $\underset{\sim}{E} > 0$, $\underset{\sim}{\varepsilon} = [(K - 1)e_{12} ; (K + 1)e_{12}]'$, but

$$\underset{\sim}{\Gamma}_E = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \neq \underset{\sim}{\Gamma}_D .$$

$\underset{\sim}{N}$ is unsymmetric:

$$\underset{\sim}{N} = e_{12}\begin{bmatrix} 5K - 2 & 2 - 2K \\ 5 - 2K & 2K - 2 \end{bmatrix},$$

and the characteristics roots must be calculated from:

$$\nu_{1,2} = \frac{e_{12}}{2}\left[(7K - 4) \pm \sqrt{23K^2 - 56K + 40}\right],$$

which are obviously not related to $\{\varepsilon_i\}$ and $\{\delta_i\}$. For example, if $K = 2$, we find

$$\nu_{1,2} = (5 \pm \sqrt{7})e_{12} = (2.35424, 7.64575) \times e_{12},$$

and

$$\underset{\sim}{A} = \begin{bmatrix} 0.33392 & 0.98467 \\ 0.94260 & 0.17441 \end{bmatrix} \; ; \; \underset{\sim}{A}^{-1} = \begin{bmatrix} -0.20049 & 1.13192 \\ 1.08355 & -0.38385 \end{bmatrix} \neq \underset{\sim}{A}' :$$

and $\underset{\sim}{A}^{-1}\underset{\sim}{N}\underset{\sim}{A}$ diagonalizes to diag (ν_1, ν_2). A typical term in the variance forecast (81) would be:

$$\underset{\sim}{X}_t\underset{\sim}{X}_t' = x_{1t}^2\begin{bmatrix} 1.14286 & 0.21429 \\ 0.21429 & 0.071429 \end{bmatrix} + x_{1t}x_{2t}\begin{bmatrix} -0.85714 & -0.28571 \\ -0.28571 & -0.42857 \end{bmatrix}$$

$$+ x_{2t}^2\begin{bmatrix} 0.28571 & 0.42857 \\ 0.42857 & 1.14286 \end{bmatrix}.$$

SUMMARY

To summarize, the linearly-dependent multinormal prior introduced in Section 6 has two important advantages over the traditional Normal-Wishart prior: it permits the specification of a much larger number of hyperparameters, which can conveniently be taken as the prior observational mean, $\underset{\sim}{m}$, and the two components of prior observational covariance, $\underset{\sim}{D}$ and $\underset{\sim}{E}$; furthermore, the prediction formula for the mean observation is of full-dimensional credibility form. Against this must be balanced the fact that the random precision is of reduced rank, leading to a prediction formula for the observational covariance that involves a new type of sufficient statistic. It will be interesting to see if this additional modelling flexibility leads to improved predictions, or if further developments of this difficult theory are possible.

BIBLIOGRAPHY

[1] Aitchison, J. and Dunsmore, I. R. (1975), STATISTICAL PREDICTION ANALYSIS, Cambridge University Press, Cambridge, (273 pp.).

[2] Ando, A. and Kaufman, G. M. (1965), "Bayesian Analysis of the Independent Multinormal Process - Neither Mean nor Precision Known," Journal of the American Statistical Association, vol. 60, pp. 347-358.

[3] Bellman, Richard (1960), INTRODUCTION TO MATRIX ANALYSIS, McGraw-Hill, New York, (328 pp.).

[4] Bühlmann, Hans (1967), "Experience Rating and Credibility," ASTIN Bulletin, vol. 4, pp. 199-207.

[5] De Groot, Morris H. (1970), OPTIMAL STATISTICAL DECISION, McGraw-Hill, New York, (489 pp.).

[6] De Vylder, Florian (1978), "Parameter Estimation in Credibility Theory," ASTIN Bulletin, vol. 10, part 1, pp. 99-112.

[7] De Vylder, Florian (1981), "Practical Credibility Theory with Emphasis on Optimal Parameter Estimation," ASTIN Bulletin, vol. 12, no. 2, pp. 115-131.

[8] De Vylder, Florian, "Practical Models - Credibility Theory," see this issue of Journal of Econometrics.

[9] Diaconis, P. and Ylvisaker, D. (1979), "Priors with Linear Posterior Expectation," Annals of Statistics, vol. 7, pp. 269-281.

[10] Dickey, J. M., Lindley, Dennis V. and Press, S. James (1981), "Bayesian Estimation of the Dispersion Matrix of a Multivariate Normal Distribution," Technical Report No. 81, Department of Statistics, University of California, Riverside.

[11] Feinberg, S. E. (1980), "Linear and Quasi-Linear Bayes Estimators," Bayesian Analysis in Econometrics and Statistics, A. Zellner (ed.), North-Holland, Amsterdam.

[12] Jewell, William S. (1974a), "Credible Means are Exact Bayesian for Simple Exponential Families," ASTIN Bulletin, vol. 8, part 1, pp. 79-90.

[13] Jewell, William S. (1974b), "Exact Multidimensional Credibility," Bulletin of the Association of Swiss Actuaries, vol. 74, part 2, pp. 193-214.

[14] Jewell, William S. (1975a), "Regularity Conditions for Exact Credibility," ASTIN Bulletin, vol. 8, part 3, pp. 336-341.

[15] Jewell, William S. (1975b), "The Use of Collaterial Data in Credibility Theory: A Hierarchical Approach," Giornale dell' Istituto Italiano degli Attuari, vol. 38, no. 1-2, pp. 1-16.

[16] Johnson, Norman L. and Kotz, Samuel (1972), DISTRIBUTIONS IN STATISTICS: CONTINUOUS MULTIVARIATE DISTRIBUTIONS, John Wiley & Sons, New York, (333 pp.).

[17] Kahn, Paul M. (ed.) (1975), CREDIBILITY-THEORY AND APPLICATIONS, Proceedings of the Berkeley Actuarial Research Conference on Credibility, September 1974, Academic Press, New York.

[18] Nobel, Ben (1969), APPLIED LINEAR ALGEBRA, Prentice-Hall, Englewood Cliffs, (523 pp.).

[19] Norberg, Ragnar (1979), "The Credibility Approach to Experience Rating," Scandinavian Actuarial Journal, vol. 1979, no. 4, pp. 181-221.

[20] Norberg, Ragnar (1980), "Empirical Bayes Credibility," Scandinavian Actuarial Journal, vol. 1980, no. 4, pp. 177-194.

[21] Norberg, Ragnar (1982), "On Optimal Parameter Estimation in Credibility," Insurance: Mathematics and Economics, vol. 1, no. 2, pp. 73-89.

[22] Press, S. James (1981), APPLIED MULTIVARIATE ANALYSIS, (Second Edition, Revised), Krieger Publishing Co., New York.

[23] Sundt, Bjørn (1981), "The Variance Part of the Credibility Premium without the Jackknife," unpublished report, Research Institute for Mathematics, ETH, Zurich, September 1981.

[24] Zehnwirth, Benjamin (1981), "The Jackknife and Variance Part of the Credibility Premium," Scandinavian Actuarial Journal, vol. 1981, no. 4, pp. 245-249.

APPENDIX

NOTATIONS, DISTRIBUTIONS, AND MOMENTS

Boldface lower case letters refer to vectors, usually p-dimensional, viz., $\underset{\sim}{x} = [x_1, x_2, \ldots, x_p]'$; boldface upper case letters refer to $p \times p$ matrices, whose elements are written in lower case, viz., $\underset{\sim}{\Omega}$ has elements $[\omega_{ij}]$. In contrast to Jewell (1974b), greek letters here refers to parameters. Random variables are indicated by a tilde over the corresponding argument; in this way we can use the usual trick of using $p(\cdot)$ for all random variables, and letting the arguments indicate the appropriate (discrete or continuous) density, so that $p(\underset{\sim}{x} \mid \underset{\sim}{\mu}, \underset{\sim}{\Omega})$ is the density of $\tilde{\underset{\sim}{x}}$, conditioned on $\tilde{\underset{\sim}{\mu}} = \underset{\sim}{\mu}$ and $\tilde{\underset{\sim}{\Omega}} = \underset{\sim}{\Omega}$. $E\{\tilde{\underset{\sim}{x}}\}$ is the usual vector mean, and $\underset{\sim}{V}\{\tilde{\underset{\sim}{x}}\} = E\{[\tilde{\underset{\sim}{x}} - E\{\tilde{\underset{\sim}{x}}\}][\tilde{\underset{\sim}{x}} - E\{\tilde{\underset{\sim}{x}}\}]'\}$ is the variance-covariance (dispersion) matrix. Sequential operators are interpreted inside-out, viz. $E\underset{\sim}{V}\{\tilde{\underset{\sim}{\mu}} \mid \tilde{\underset{\sim}{\Omega}}\}$ means: first the dispersion of $\tilde{\underset{\sim}{\mu}}$, conditional upon $\tilde{\underset{\sim}{\Omega}} = \underset{\sim}{\Omega}$; then the expectation over all values of $\tilde{\underset{\sim}{\Omega}}$.

Our definition of distributions is taken mostly from Johnson and Kotz (1972) with the exception that we always emphasize the precision (parameter) matrix, instead of the dispersion.

The *multivariate-normal (multinormal)* distribution with mean vector $\underset{\sim}{m}$ and precision-matrix $\underset{\sim}{W}$ has $\tilde{\underset{\sim}{x}} \sim N_p(\underset{\sim}{m}; \underset{\sim}{W})$ if

$$p(\underset{\sim}{x}) = (2\pi)^{-p/2} |\underset{\sim}{W}|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}(\underset{\sim}{x} - \underset{\sim}{m})'\underset{\sim}{W}(\underset{\sim}{x} - \underset{\sim}{m})\right\} , \qquad (\underset{\sim}{x} \in R^p) .$$

Moments are:

$$E\{\tilde{\underset{\sim}{x}}\} = \underset{\sim}{m} \quad ; \quad \underset{\sim}{V}\{\tilde{\underset{\sim}{x}}\} = \underset{\sim}{W}^{-1} .$$

The *multivariate* Student-t ("with common denominator") distribution with α degrees of freedom, mean vector $\underset{\sim}{m}$, and (symmetric) precision-parameter matrix $\underset{\sim}{W}$ has $\tilde{\underset{\sim}{x}} \sim S_p(\alpha;\underset{\sim}{m};\underset{\sim}{W})$ if:

$$p(\underset{\sim}{x}) = \frac{\Gamma\left(\frac{1}{2}(\alpha + p)\right)}{(\pi\alpha)^{p/2}\Gamma\left(\frac{\alpha}{2}\right)} |\underset{\sim}{W}|^{\frac{1}{2}}[1 + \alpha^{-1}(\underset{\sim}{x} - \underset{\sim}{m})'W(\underset{\sim}{x} - \underset{\sim}{m})]^{-\frac{1}{2}(\alpha+p)} , \qquad (\underset{\sim}{x} \in R^p) .$$

Moments are:

$$E\{\tilde{\underset{\sim}{x}}\} = \underset{\sim}{m} \quad ; \quad \underset{\sim}{V}\{\tilde{\underset{\sim}{x}}\} = \frac{\alpha}{\alpha - 2}\underset{\sim}{W}^{-1} .$$

The *Wishart* distribution with α degrees of freedom and a (symmetric) precision-parameter matrix $\underset{\sim}{W}$ is defined only for a random, *symmetric* matrix $\tilde{\underset{\sim}{\Omega}}$ over values in $R^{\frac{1}{2}p(p+1)}$ that make it positive definite. $\tilde{\underset{\sim}{\Omega}} \sim W_p(\alpha,\underset{\sim}{W})$ if the density is:

$$p(\underset{\sim}{\Omega}) = \left[2^{\frac{1}{2}\alpha p}\Gamma_p\left(\frac{1}{2}\alpha\right)\right]^{-1} |\underset{\sim}{W}|^{\frac{1}{2}\alpha}|\underset{\sim}{\Omega}|^{\frac{1}{2}(\alpha-p-1)} \exp\left[-\frac{1}{2}\operatorname{tr}(\underset{\sim}{\Omega}\underset{\sim}{W})\right] , \qquad (\underset{\sim}{\Omega} > 0) .$$

Letting $\tilde{\underset{\sim}{\Omega}}^{-1} = \tilde{\underset{\sim}{\Sigma}} = (\tilde{\sigma}_{ij})$ (the usual variance-covariance matrix), we have

$$E\{\tilde{\underset{\sim}{\Omega}}\} = \alpha\underset{\sim}{W}^{-1} \quad ; \quad E\left\{\tilde{\underset{\sim}{\Sigma}}\right\} = (\alpha - p - 1)^{-1}\underset{\sim}{W} ;$$

and the covariances are: (see Press (1981))

$$C\{\tilde{\omega}_{ij};\tilde{\omega}_{kl}\} = \alpha(w^{ik}w^{jl} + w^{il}w^{jk}) ;$$

$$C\{\tilde{\sigma}_{ij};\tilde{\sigma}_{kl}\} = [(\alpha - p)(\alpha - p - 1)(\alpha - p - 3)]^{-1}\left[2(\alpha - p - 1)^{-1}w_{ij}w_{kl} + w_{ik}w_{jl} + w_{il}w_{kj}\right]$$

for *all* (i,j,k,l) , where $w^{ij} = (\underset{\sim}{W}^{-1})_{ij}$.

The *Gamma* distribution with shape parameter γ and scale parameter β has $\tilde{\omega} \sim G(\gamma,\beta)$ if:

$$p(\omega) = \frac{\beta(\beta\omega)^{\gamma-1}e^{-\beta\omega}}{\Gamma(\gamma)} , \qquad (\omega > 0) .$$

Moments are:

$$E\{\tilde{\omega}\} = \gamma\beta^{-1} \quad ; \quad E\{\tilde{\omega}^{-1}\} = (\gamma - 1)^{-1}\beta \ ;$$

$$V\{\tilde{\omega}\} = \gamma\beta^{-2} \quad ; \quad V\{\omega^{-1}\} = (\gamma - 1)^{-2}(\gamma - 2)\beta^{2} .$$

It follows that $W_1(\alpha,w_{11}) = G\left(\frac{1}{2}\alpha,\frac{1}{2}w_{11}\right)$ is a Chi-squared density.

DATE
ILME

got from $\underset{\sim}{M} = \underset{\sim}{D}^{-1}\underset{\sim}{W}^{-1}$, with $\underset{\sim}{D} = \underset{\sim}{V}(\tilde{\underset{\sim}{u}})$ now an arbitrary, symmetric positive definite matrix to be estimated. In other words, apart from $\underset{\sim}{m}$ and $\underset{\sim}{C} = \underset{\sim}{V}(\tilde{\underset{\sim}{x}})$

then compute $\underset{\sim}{N} = \underset{\sim}{E}\underset{\sim}{D}^{-1}$, $\underset{\sim}{x}_0 = \underset{\sim}{N}\underset{\sim}{m}$, and $\underset{\sim}{U}_0 = (\underset{\sim}{N} + \underset{\sim}{I})\underset{\sim}{E}$. The eigenvalues of $\underset{\sim}{N}$ are then computed, and if these are negative (or less than unity), then the assumptions about $\underset{\sim}{E}$, $\underset{\sim}{D}$ must be inconsistent with the LDMEF model. Usually, however, there will be no difficulty at this point, and $\underset{\sim}{A}$ is

$$E(\tilde{\underset{\sim}{x}}) = \underset{\sim}{m} \; ; \; \underset{\sim}{V}(\tilde{\underset{\sim}{x}}) = \underset{\sim}{W}^{-1} .$$

$$C(\bar{\sigma}_{ij};\bar{\sigma}_{kl}) = [(a-p)(a-p-1)(a-p-3)]^{-1}\left[2(a-p-1)^{-1}w_{ij}w_{kl} + w_{ik}w_{jl} + w_{il}w_{kj}\right] .$$

for *all* (i,j,k,l) , where $w^{ij} = (\underset{\sim}{W}^{-1})_{ij}$.